Response of Solar Irradiance to Sunspot-area Variations — **Source link**

T. Dudok de Wit, Greg Kopp, Alexander Shapiro, V. Witzke ...+1 more authors

Institutions: Max Planck Society

Related papers:

- Solar irradiance variations induced by faculae and small magnetic elements in the photosphere

- Readdressing the UV solar variability with SATIRE-S: non-LTE effects

- Using Precise Solar Limb Shape Measurements to Study the Solar Cycle

- Solar irradiance modulation by active regions during 1980

- Towards Long-Term Solar Irradiance Modelling: Network Contrasts from Magneto-Convection Simulations

Response of Solar Irradiance to Sunspot-area Variations

Thierry Dudok de Wit, G. Kopp, A. Shapiro, V. Witzke, Matthieu Kretzschmar

CrossMark

Response of Solar Irradiance to Sunspot-area Variations

T. Dudok de Wit[1] , G. Kopp[2,3], A. Shapiro[3], V. Witzke[3], and M. Kretzschmar[1]
[1] LPC2E, University of Orléans and CNRS, France
[2] Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, Colorado, USA
[3] Max-Planck-Institut für Sonnensystemforschung, Göttingen, Germany

Abstract

One of the important open questions in solar irradiance studies is whether long-term variability (i.e., on timescales of years and beyond) can be reconstructed by means of models that describe short-term variability (i.e., days) using solar proxies as inputs. Preminger & Walton showed that the relationship between spectral solar irradiance and proxies of magnetic-flux emergence, such as the daily sunspot area, can be described in the framework of linear system theory by means of the impulse response. We significantly refine that empirical model by removing spurious solar-rotational effects and by including an additional term that captures long-term variations. Our results show that long-term variability cannot be reconstructed from the short-term response of the spectral irradiance, which questions the extension of solar proxy models to these timescales. In addition, we find that the solar response is nonlinear in a way that cannot be corrected simply by applying a rescaling to a sunspot area.

Key words: methods: data analysis – Sun: activity – Sun: magnetic fields – Sun: UV radiation

1. Introduction

The variability of solar irradiance was unambiguously proven almost four decades ago, after the launch of the *NIMBUS* 7 mission (Hickey et al. 1980). Even prior to this, it was suspected that a major contributor to climate change could be caused by a variable Sun (Eddy 1976), after which studies of solar irradiance variability attracted considerable attention. Over the last few years significant progress has been achieved in understanding the physical causes of solar variability and the mechanisms by which it can influence Earth's climate. The mechanisms causing such irradiance variations have largely been identified (Domingo et al. 2009; Solanki et al. 2013; Yeo et al. 2017), with magnetic-field evolution and solar rotation causing variations on timescales larger than a day, while granulation and oscillations are responsible for variations on shorter timescales.

An uninterrupted series of space-based missions, beginning with *NIMBUS* 7, have monitored the variability of the total solar irradiance (TSI), which is the spectrally integrated solar-radiative flux normalized to 1 au from the Sun, with high precision and accuracy (Kopp 2016; Dudok de Wit et al. 2017). Current TSI models can reproduce these measurements on timescales from minutes to decades (Seleznyov et al. 2011; Yeo et al. 2014; Coddington et al. 2016; Shapiro et al. 2017). Daily measurements of the spectral solar irradiance (SSI) in the ultraviolet began at about the same time and have since been extended to span the visible and near-infrared up to 2.4 μm.

However, a number of important open questions remain. In particular, while models and observations of the variability of the SSI agree reasonably well in the short term (i.e., on timescales from a few days to solar-rotational periods; Ermolli et al. 2013; Solanki et al. 2013), the amplitude and even the phase of the variations on timescales of the 11-year solar-activity cycle and longer are still unclear (Harder et al. 2009; Wehrli et al. 2013). Knowledge of such long-term SSI variations is crucial for assessing the impact of solar variability on the Earth's climate. The lack of observational constraints on their magnitude has unfortunately hampered Sun-climate studies (e.g., Oberländer et al. 2012; Ball et al. 2016).

In order to solve this problem, numerous attempts have been made to extrapolate SSI variations to multi-decadal timescales by means of empirical models that reproduce the dynamics at shorter timescales (Lean 2000; Preminger & Walton 2006b; Coddington et al. 2016). Different physical mechanisms may affect the SSI variability on different timescales, however (Shapiro et al. 2015), making such extrapolations non-trivial. Nevertheless, such irradiance models appear to reproduce the observations reasonably well in the short term, at least over the four decades for which direct observations are available. Their ability to properly reproduce longer-term variations is an open and hotly debated question that remains unanswered partly because of a lack of stable, long-term SSI measurements, particularly in the visible and near-infrared spectral bands where relative solar-variability is small.

Irradiance-reconstruction models rely on several assumptions. First, they assume that the SSI can be reconstructed from a linear or weakly nonlinear combination of one or more solar proxies such as sunspot number or sunspot area. A second assumption is that the relation between the SSI and proxies is time-invariant, which is needed in order to reconstruct the SSI backward in time before the observations began. However, the models are determined using data from only the relatively short overlap period (typically one to four decades) during which both the SSI observations and proxies are available, and thus rely on time-invariance for extrapolations over their multi-century pre-observation time ranges. A third assumption, which is an important focus of this article, is the instantaneity of the response of the SSI to solar proxies, particularly to flux emergence as indicated by changes in the sunspot area. Preminger & Walton (2005, 2006a) were among the first to consider the convolutive nature of this particular response and the possibility of a time delay relative to changes in the proxies to which they are causally related. Incorporating such a delay improves the impulse-response models' reconstruction capacity and also gives deeper insight into their physical interpretation.

The usual framework for describing such a non-instantaneous response between an input (a solar proxy) and an output (the SSI) is that of linear systems with transfer functions that

are classically represented by their impulse response (Antsaklis & Michel 2005). The impulse response relating the SSI to a solar proxy such as the sunspot area indicates by how much an increase in sunspot area leads to an excess or deficit of spectral irradiance. Such relations may help determine whether solar brightness variations are spot- or faculae-dominated, which in turn helps explain solar variability on longer timescales. Several recent studies have been devoted to this question (Preminger et al. 2011; Woods et al. 2015; Shapiro et al. 2016).

Motivated by this important question, here we evaluate the impulse response of the SSI to the sunspot area along the line initiated by Preminger & Walton (2005, 2006a). However, we carry the study further by (1) eliminating solar-rotational effects that hamper the proper estimation of the impulse response and its connection to the energy budget, and (2) introducing a more general transfer function model that allows the separation of causally related effects from other contributions. Our results reveal the nonlinearity of the impulse response during the phase when there is flux emergence (i.e., before the sunspot decays) and reveal that it not possible to reconstruct long-term variability from a linear impulse-response model based on short-term measurements. We stress that our objective is to infer physical information from the impulse response and not to find the best model for reconstructing the SSI from sunspots, which is a different problem.

The paper is organized as follows. The methodology is described in Section 2, followed by a description of the data sets in Section 3. We discuss the results in Section 4, and summarize our conclusions in Section 5.

2. Methodology

2.1. Impulse Response with No Residual

Our approach is conceptually similar to that introduced by Preminger & Walton (2005, 2006a, 2006b, 2007; henceforth denoted as PW), with two major differences that are described in Sections 2.2 and 2.3. The studies by PW suggested the possibility of describing the relationship between variations in the sunspot area and in the SSI by means of a linear time-invariant system. An essential feature of such time-invariant systems is that they are entirely characterized by their impulse response $h(t)$, which relates their response $y(t)$ to a given stimulus or input $u(t)$ by

$$y(t) = h(t) * u(t) = \int_{-\infty}^{\infty} h(t - \tau) \, u(\tau) \, d\tau. \quad (1)$$

The Fourier transform turns this convolution into a product:

$$Y(\omega) = H(\omega)U(\omega), \quad (2)$$

where $Y(\omega)$ stands for the Fourier transform of $y(t)$, etc. Once we have estimated the impulse response $h(t)$ or the transfer function $H(\omega)$, we can model the response to any type of temporal evolution of the input. The impulse response may be estimated by working in Fourier space: for each frequency ω we then perform a linear regression of $Y(\omega)$ versus $U(\omega)$.

To illustrate this concept let us imagine a simple dynamical system that is described by the linear differential equation

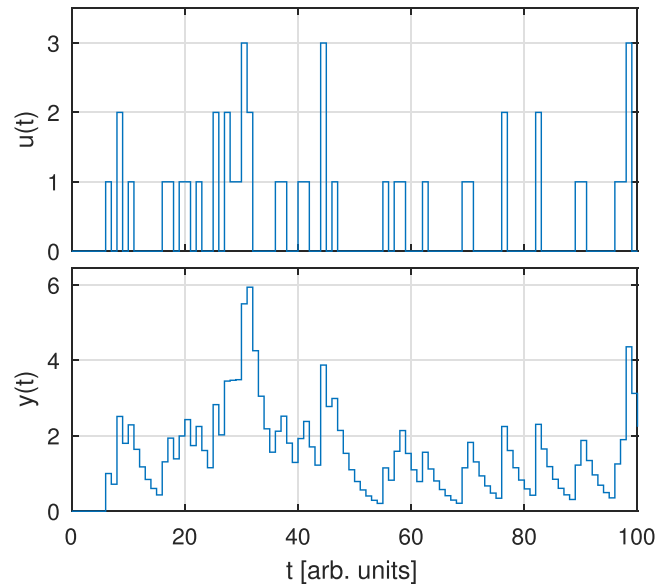$$\frac{d}{dt} y(t) = u(t) - \frac{y(t)}{\tau}. \quad (3)$$

Figure 1. Example showing an input $u(t)$ that consists of random pulses, and the associated response $y(t)$ for the model described by Equation (3) with a relaxation time $\tau = 4$.
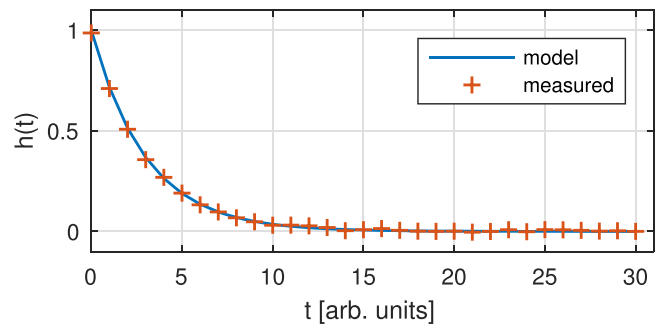

Figure 2. Illustration of the excellent agreement between the theoretical impulse response $h(t)$ for the model described by Equation (3) (blue line) and that obtained by computing the inverse Fourier transform of $H(\omega) = Y(\omega)/U(\omega)$ (red plus signs) using the data from the example shown in Figure 1 after adding noise to them.

For a discrete time series this first-order differential equation can be reproduced by a simple first-order autoregressive model with exogenous inputs (Ljung 1997). As an example, Figure 1 shows an input $u(t)$ that consists of a series of discrete random pulses; these could be, for instance, a time series of the sunspot number near solar minimum. The shown modeled output $y(t)$ computed from Equation (3) tracks the input with a relaxation decay time τ. This output is qualitatively similar to what we shall observe later for that of the UV irradiance to changes in sunspot area.

Taking the Fourier transforms of the input and output from the example in Figure 1, we estimate the impulse response $h(t)$, which is illustrated in Figure 2 together with its theoretical model based on Equation (3).

We simulated the dynamical system in this example using conditions that are similar to the ones encountered with real solar data, such as the length of the data sequences and the inclusion of ∼20% additive white noise in the input and output. As shown in Figure 2, the reconstructed impulse response agrees remarkably well with the theoretical one, in spite of the contamination of the data by noise. The estimation of such

impulse responses from finite and noisy records is a central task in the field of system identification (Ljung 1997) and a considerable amount of research has been devoted to the design of robust estimators, with many applications in automatic control, signal processing, telecommunications, etc. Here, our main objective is to estimate the impulse response from the response of the SSI to changes in sunspot area in order to better understand the nature of SSI variations.

2.2. Impulse Response with Residual

The impulse response model we have described is one that is commonly utilized in linear system theory (Antsaklis & Michel 2005). However, the variability of solar irradiance may have different drivers, including some that are not directly caused by a single selected input $u(t)$. In other words, some part of the SSI variations may arise from processes independent of sunspots and products of their decay. To account for this, we extend the definition of the dynamical response by allowing two contributions: (1) the classical impulse response that relates output to input; and (2) an additional (but not necessarily small) residual contribution $r(t)$ that captures whatever is not described by the classical linear impulse response. Note that we do not impose any model for this residual, which merely describes what remains after modeling the convolution of a stimulus with the impulse response. Instead of Equation (1), we now have

$$y(t) = h(t) * u(t) + r(t), \qquad (4)$$

whose Fourier transform gives

$$Y(\omega) = H(\omega)U(\omega) + R(\omega). \qquad (5)$$

This system is fully characterized by $h(t)$ and $r(t)$. Here again, it is more convenient to work in Fourier space. In practice, for each frequency ω, we have an ensemble of M estimates of $\{U_k(\omega)\}$ and $\{Y_k(\omega)\}$, with $k = 1, 2, \ldots, M$. The problem then amounts to finding for each frequency the two coefficients $\{H(\omega), R(\omega)\}$ that solve the overdetermined set of equations $Y_k(\omega) = H(\omega)U_k(\omega) + R(\omega)$. This is typically done by total least squares, as classical least squares are not suitable because both Y and U have errors. The impulse response $h(t)$ and the residual $r(t)$ are then obtained by the inverse Fourier transform of the slope and the intercept of these lines, respectively.

To illustrate the importance of the residual term in Equation (4), we added an artificial trend to the output shown in Figure 1 and estimated the resulting impulse response $h(t)$ together with the residual $r(t)$, again after adding noise to the input and output. Figure 3 compares the scatterplot of the output $|Y(\omega)|$ versus the input $|U(\omega)|$ at a given frequency with and without the added trend. When there is no trend in the data, the linear transformation intercepts the origin, as expected. However, this intercept is nonzero when the output has an additional trend. We thus have a simple criterion for both testing the presence of a residual and estimating it.

Figure 3 also compares the reconstructed determination of the artificial trend to what was actually applied, and confirms the ability of our methodology to properly extract both $h(t)$ and $r(t)$ from the observations, even when the latter are corrupted by noise. If no residual is allowed for in the model when there is one in the observations, then the impulse response will be biased and incorrect. In previous studies of the transfer function applied to solar irradiances, no such residual term was
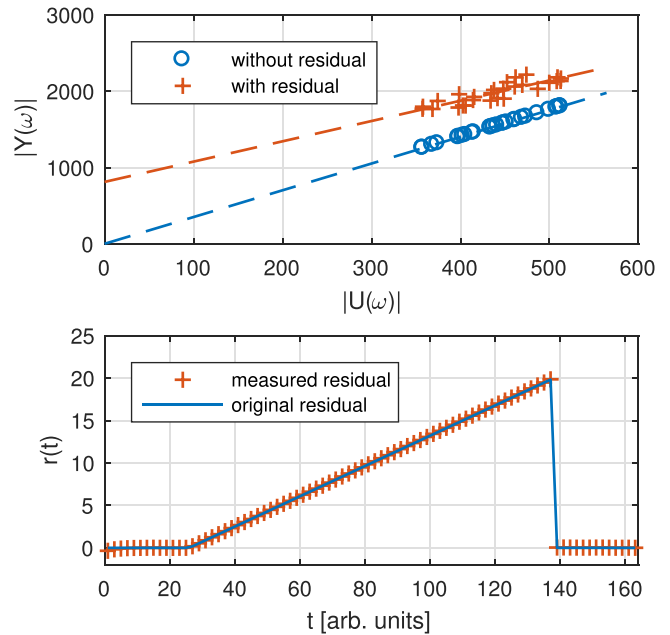

Figure 3. Upper plot: representation of $|Y(\omega)|$ vs. $|U(\omega)|$ at a particular frequency for the example shown in Figure 1 with and without an added artificial trend to the input time series. The regression is done in complex space; here, we display the modulus for visualization purposes only. Also shown is the linear fit for each case, from which the impulse function and the residual can be estimated. Lower plot: the reconstructed residual (red crosses) closely matches the original trend (blue line) that was artificially added to the output.

included, potentially biasing the determination of the impulse response by background residuals. We find that the inclusion of this residual term affects the resulting solar-response determinations, therefore we systematically include it in our model.

System identification often relies on parametric approaches to estimate impulse responses because of their better noise immunity (Ljung 1997). However, such approaches require the explicit design of a model of the response. While the impulse response can generally be described by a simple parametric model (e.g., an exponential in the example above), this is not true of the residual, whose variation in time can be arbitrarily complex. For that reason, parametric models are not appropriate for determinations of solar irradiance responses to various sources, as the measurements of each can include background effects that may be due to non-solar (i.e., instrumental) causes. We thus utilize a more general non-parametric (Fourier) approach (e.g., Schoukens et al. 2009).

2.3. "Snapshots" Remove Solar-rotational Variability

Aside from the inclusion of the residual term in Equation (4), there is a second important difference between our approach and that of PW. In the latter, all inputs and outputs used daily cadence time-series data directly, thus incorporating both the desired solar-dynamical effects as well as unrelated solar-rotational effects (such as center-to-limb variations of contrasts of solar active regions as well as foreshortening). These solar-rotational effects alter the true impulse-response function by modulating it with a 27-day oscillation, which, in addition, is wavelength-dependent. Although we could in principle deconvolve the two effects, it is considerably easier to reorganize the data in a way that eliminates the rotational modulation. To do so, we consider a stroboscopic approach, taking a "snapshot" of

the Sun once every solar rotation. Observing the Sun at this cadence allows us to concentrate on the dynamical evolution of active regions while strongly reducing the solar-rotational influences on their visibility; we thus separate most of the solar-dynamical response from the rotationally induced variability. The tradeoff with this approach is coarser time resolution, since we cannot resolve timescales shorter than one solar rotation.

To illustrate this stroboscopic approach, consider a sequence of daily values $\{y(t_0), y(t_0 + 1), y(t_0 + 2), \ldots\}$. We extract from this a new record $\{y(t_0), y(t_0 + 27), y(t_0 + 54), \ldots\}$ that is now sampled every 27 days. One may also consider the sequence that is observed one day later: $\{y(t_0 + 1), y(t_0 + 28), y(t_0 + 55), \ldots\}$, and likewise build additional records. In total, 27 different records can thus be generated.

In this paper we analyze solar irradiance and solar proxy data from ∼12 years of observations (4401 daily values), giving 27 records of 163 time steps for each, with each time step corresponding to a different solar rotation. These realizations are not fully independent; however, they contain exactly the same amount of information as the original time-series records. Let us therefore stress that we exploit the full data set, with no addition nor loss of information.

Fortunately, because our 12 years of data correspond to nearly one solar cycle, each stroboscopic record is close to periodic. This avoids the need to apply cumbersome windowing procedures that might otherwise be required before applying Fourier transforms.

The strength of the impulse-response approach we utilize lies in its ability to isolate and quantify the response of the Sun to changes in net disk-integrated sunspot area by observing the cumulative effect of an arbitrarily large number of spots for a sufficiently long time. Let us stress that we relate variations in the irradiance to variations in the sunspot area without distinguishing between short-lived and long-lived spots. In particular, a newly emerging sunspot whose area exactly compensates a decaying sunspot will cause no variation in the irradiance. This is not exactly true, as the contribution to the SSI of features such as faculae is position-dependent. By using disk-integrated proxies we force the impulse response to describe an average response of the Sun. An important condition for this is the linearity of the solar response, which will be addressed in Section 4.5. The main consequence of the location-dependent effects of the sunspots, combined with their varying locations of emergence on the solar disk, will be a larger uncertainty in the values of the impulse response.

3. Data

3.1. The Solar Indices Used

For solar irradiances, we utilize daily observations made by the *Solar Radiation and Climate Experiment* (*SORCE*) mission (Rottman 2005) from 115.5 to 2000 nm, and observations by the *Thermosphere Ionosphere Mesosphere Energetics and Dynamics* (*TIMED*) mission (Woods et al. 2005) from 29.5 to 115.5 nm. These two missions provide the longest time-interval for which the SSI was continuously monitored from the UV to the near-infrared, namely from 2003 May 15 to 2015 June 14. The spectral resolution is 1 nm from 29.5 to 310 nm and then progressively increases to 15.7 nm at longer wavelengths. For convenience, we concentrate on the irradiance integrated in six spectral bands: the extreme-ultraviolet (EUV) from 30 to 121 nm; the far-UV (FUV) from 122 to 200 nm; the middle-UV

(MUV) from 200 to 300 nm; the near-UV (NUV) from 300 to 400 nm; the visible (VIS) from 400 to 600 nm; and the near-infrared (NIR) from 600 to 2000 nm. The SSI data contain a small fraction of observational gaps or outliers that are filled in by expectation-maximization (Dudok de Wit 2011). This filling considerably facilitates the Fourier analysis and has no significant impact on the results.

The input to our solar irradiance impulse-response model should be a time series related to the flux emergence of active regions on the visible side of the solar disk, as these regions are known to be causally related to the irradiance. In addition to the irradiances above, we consider five observables that provide inputs from which the irradiance is modeled: the daily sunspot area (DSA) on the solar disk provided by the Royal Greenwich Observatory; the solar radio flux at 10.7 cm (F10.7) from the Penticton Observatory; the solar radio flux at 30 cm (F30) from the Nobeyama Observatory; the core-to-wing ratio (Mg II) of the Mg II line from the University of Bremen; and the TSI from *SORCE*/TIM. The latter, being from a dedicated TSI instrument, is independent of the SSI provided by other instruments on the *SORCE* mission. Because the solar irradiance and these five observables are all disk-integrated measurements, the impulse function determined from each observable is the average response of the solar-radiative output without regard to the location(s) of the responsible active area(s) on the solar disk.

PW showed that an ideal candidate for the input time series is the DSA, which is given as a percentage of the solar hemisphere. For convenience, we divide the official DSA by 100, so that its units are 10^{-4} of the fractional surface of the solar hemisphere (i.e., 100 μhem). In these units, a new sunspot typically corresponds to an increase in $u(t)$ of the order of one unit; high levels of solar activity typically correspond to values of 10 and above.

For proper determination of the impulse response, it is important to consider the response of the Sun to a varying input and thus ignore any constant background or baseline. A natural choice for the latter is the level of activity corresponding to a quiet Sun. For the SSI we define the response as $y(t) = \text{SSI}(\lambda, t) - \text{SSI}_{\text{QS}}(\lambda)$, where $\text{SSI}_{\text{QS}}(\lambda)$ is the quiet Sun level that is estimated from periods having at least 30 consecutive days with no sunspots during the prolonged solar minimum of 2008–2009. For the DSA, the zero level is the natural baseline.

3.2. The Need for Accommodating Residual Contributions

As shown in Figure 3, the scatterplot of $Y(\omega)$ versus $U(\omega)$ can clearly indicate whether a purely linear impulse-response model is applicable and the need for the inclusion of a residual trend. Figure 4 similarly illustrates this using the DSA as input and the Mg II index as output. The nearly linear relationship between the two quantities supports the validity of our linear and time-invariant impulse-response model. More importantly, it reveals the presence of a large offset $R(\omega)$ that cannot be neglected, which confirms the importance of including a residual term in our model. While Figure 4 illustrates the linear relationship for the lowest frequency only, qualitatively similar results are obtained at all other frequencies. Except for the lowermost frequencies, however, the residual trend contains little power and the noise level is relatively high, which makes the linear relationship, and in particular the offset, less apparent.
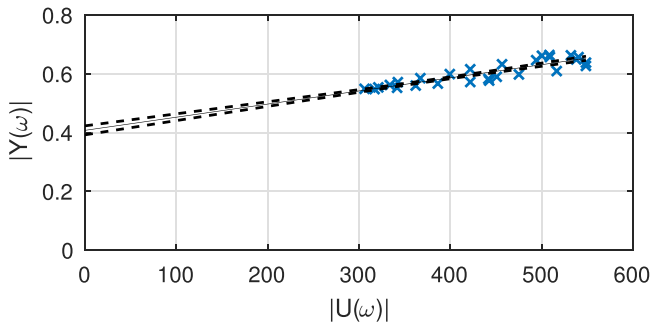
Figure 4. Scatterplot of the lowest nonzero frequency (4401^{-1} day^{-1}) of the Fourier component of the Mg II index ($|Y(\omega)|$) vs. that of the DSA ($|U(\omega)|$). Here, we display the modulus solely for visualization purposes; the true regression is done in complex space. Also shown is the linear fit, from which the impulse function and the residual are estimated. The dashed lines correspond to $\pm$ one standard deviation of that fit.

4. Results and Discussion

4.1. Impulse Responses of Solar Indices to Sunspots

The central results of our study are the impulse responses $h(t)$ of the Sun to temporal changes in net sunspot area, which are representative of varying solar-surface magnetic activity.

These impulse-response results are plotted in Figure 5 for the SSI over different spectral bands and in Figure 6 for the other solar observables. Each quantity's impulse response can be interpreted as the average variation in that quantity when the DSA increases by one unit (i.e., 100 μhem) for one solar rotation and then returns to its previous level for subsequent rotations. Although $h(t)$ is estimated for the full time span (i.e., 163 solar rotations), we display it only for the first 20 solar rotations, of which only the first few are significantly nonzero.

Figure 5 illustrates that an increase in sunspot area on average initially generates an excess of emission in the shorter-wavelength UV bands and a deficit in the NUV and longer-wavelength bands, as indicated by the respectively positive (excess) and negative (deficit) values of $h(t)$ at $t = 0$ in the impulse responses; Figure 6 shows a similar initial deficit for the TSI. The deficits are due to sunspot-darkening initially exceeding the contributions from bright structures such as faculae and plages at these wavelengths, despite those bright structures generally appearing along with newly emerging or increasing-size sunspots. We find these negative values occur only at $t = 0$, meaning facular brightening overwhelms sunspot-darkening within one solar-rotation period.

Although this sunspot-darkening is a known result (e.g., Spruit 2000; Fröhlich & Lean 2004; Foukal 2013), our approach shows it in a more compact and quantitative way: $h(t)$ quantitatively describes the average change in measured SSI for a unit increase in sunspot area, and the method of using 27-day "snapshots" avoids the need for separately deconvolving solar-rotational effects.

The impulse response for subsequent solar rotations ($t > 0$) describes the decay of the solar response to flux emergence after the sunspot area has returned to its pre-emergent value. We find that all spectral bands in Figure 5 exhibit an excess of radiation after the initial $t = 0$ response. This excess decays with an average time of 3.0 ± 0.7 solar rotations and is associated with the gradual conversion of remnants of the active region into enhanced magnetic network.
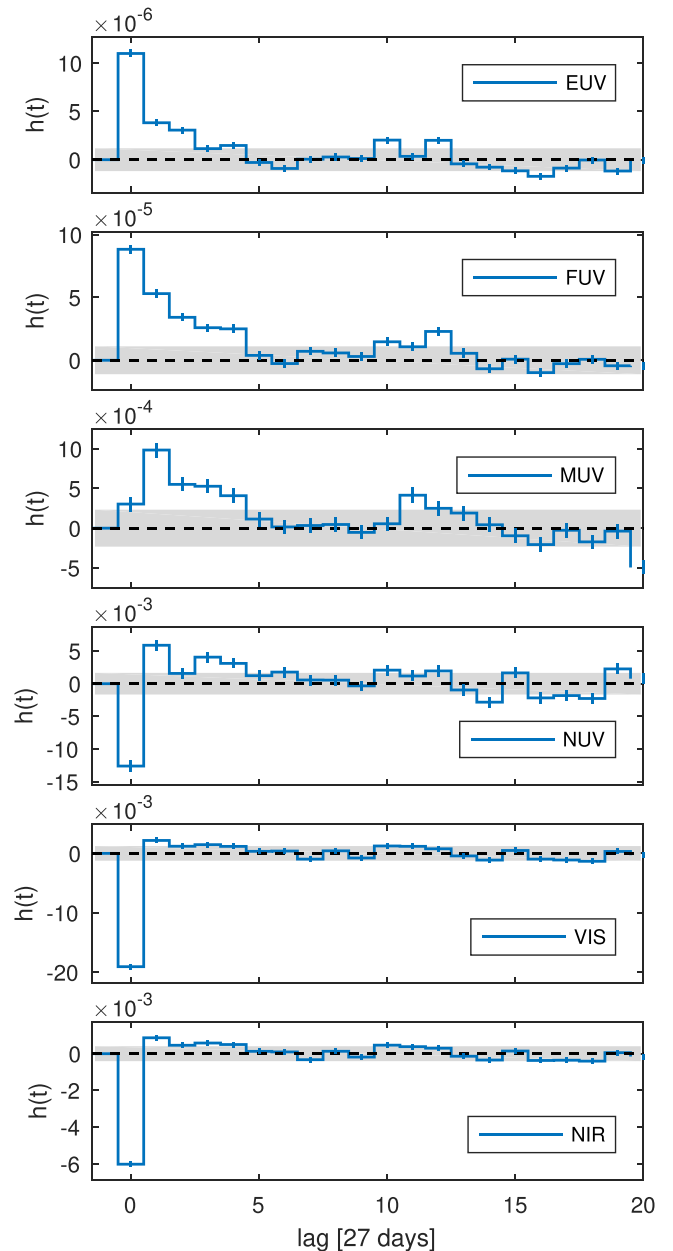

Figure 5. The impulse response of the spectral solar irradiance due to the emergence of sunspots is plotted for six spectral bands. $h(t)$ is expressed in W m^{-2} per unit increase in the DSA, which is in 100 μhem. The shaded bands denote amplitudes for which the impulse response cannot be meaningfully distinguished from a randomly varying input. The width of this band, as well as the error bars plotted, correspond to $\pm$ one standard deviation.

The decay can be tracked at best for 3–4 rotations before the remnants of the active region are buried in the noise floor, which is represented in Figure 5 by a shaded band. This noise floor is obtained from surrogate data (Schreiber & Schmitz 2000): we generate a large ensemble of synthetic records that have the same spectral properties and probability density as the original input and output but have randomized phases in Fourier space and thus are not causally related. From the standard deviation of samples of 600 impulse responses we obtain an indication of the level below which $h(t)$ cannot be meaningfully interpreted. With longer records and better
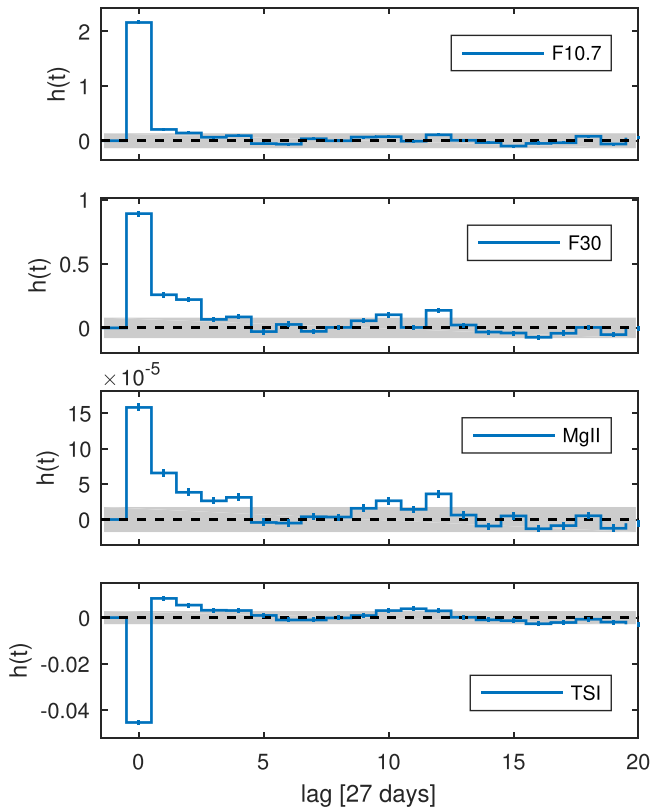
Figure 6. The impulse responses due to sunspot emergence are plotted for various solar quantities: the radio flux at 10.7 cm (F10.7) and at 30 cm (F30); the Mg II core-to-wing index; and the total solar irradiance (TSI). The shaded band and the errors again indicate ± one standard deviation.

statistics, this noise level shrinks and the decay can be tracked for longer periods; with the F10.7 index (for which observations started in 1947), for example, up to seven rotations can be observed before reaching the noise floor.

The number of nonzero terms in the impulse response is indicative of the number of solar rotations required for the signature in the SSI due to a changing sunspot area to vanish. Since only the 4–6 first terms of the impulse response are significantly nonzero, we conclude that the SSI has little or no memory of DSA variations by that same number of solar rotations. That is, the SSI cannot be delayed with respect to the DSA by more than a few months. We find no evidence for lags of several years or completely out-of-phase responses of the SSI, as suggested for example by Woods et al. (2015).

The shape of the impulse response in Figure 6 helps clarify the behavior of solar proxies. The 30 cm radio flux (or F30 index), for example, has recently been advocated as a better proxy than the F10.7 for the UV (Dudok de Wit et al. 2014) because it contains a stronger contribution from Bremsstrahlung, which is associated with UV-emitting bright loops above active regions. Figure 6 shows that both F10.7 and F30 have a bright peak at $t = 0$ (which mainly comes from gyro-emissions that are associated with the active regions) followed by a decay that is considerably brighter in F30 than in F10.7. A comparison of the impulse responses shows that the variability of the F30 index is indeed considerably closer to that of the Mg II index than to that of the F10.7 index.
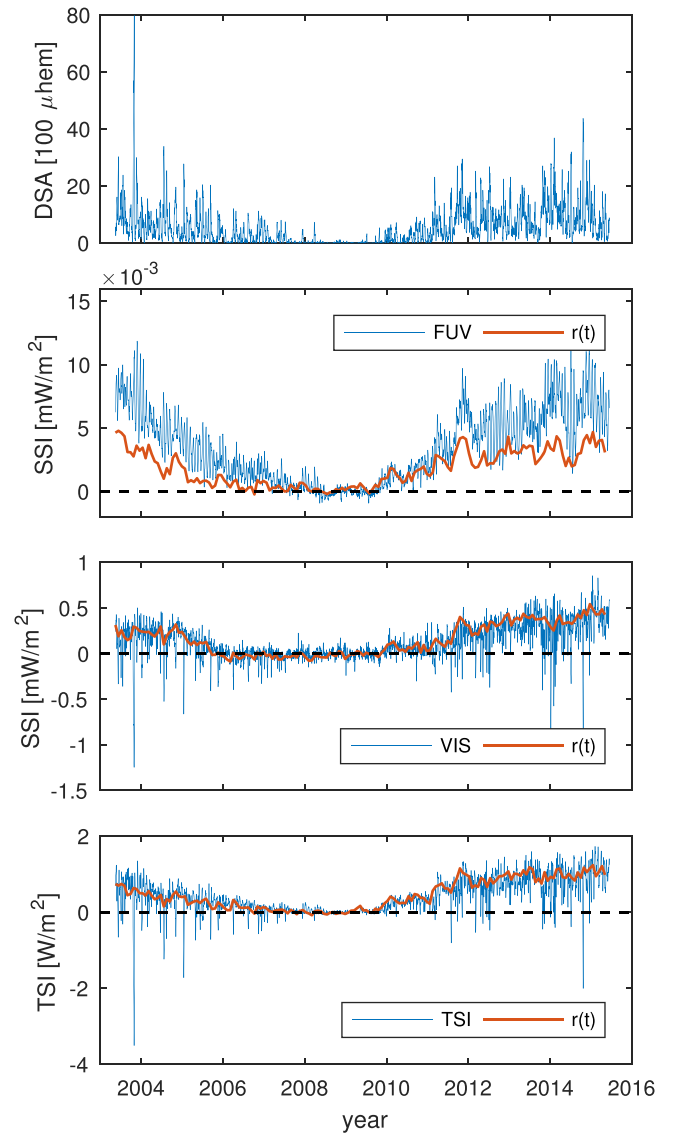

Figure 7. The input DSA (upper plot) is modeled to three outputs, with the observations (blue) and their residual contributions $r(t)$ (red) shown: the solar irradiance in the FUV (second plot); the solar irradiance in the visible (third plot); and the TSI (fourth plot). The quiet Sun level has been subtracted from each quantity. The 1σ uncertainty on $r(t)$ is approximately one-eighth of its amplitude.

4.2. The Residual Contribution

The residual contribution $r(t)$ (see Equation (2)) describes variations that cannot be reproduced by the time-invariant impulse response. Most of the spectral power of $r(t)$ is concentrated in low frequencies, primarily capturing slow variations on timescales of several months and longer. Note that the slow variation of the residual contribution is an observational result and is not imposed by our transfer function model. Figure 3 demonstrates that $r(t)$ can describe sharp discontinuities as well.

Figure 7 illustrates the residual contribution $r(t)$ for two typical cases: in the FUV, where it captures approximately half of the long-term variability, and in the visible, where it accounts for almost all of the long-term variability. The relatively large amplitude of $r(t)$ is an important finding, since it means that a significant portion of the variability in the irradiance measurements cannot be reconstructed from changes
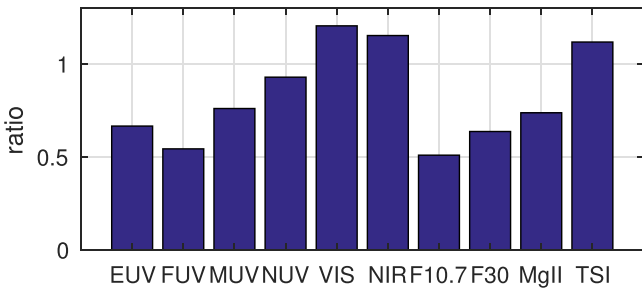
Figure 8. Ratio between the amplitudes of $r(t)$ and $y(t)$ after decimating the latter to 27 days. Values larger than one are possible because we use the standard deviation as a coarse measure of amplitude.

in the sunspot area alone (at least, not with a linear model). This questions the validity of models that reconstruct the SSI from one single solar proxy, such as the net sunspot area or the even more-limited sunspot number; or it questions the long-term stability of the measurements themselves.

A nonzero residual contribution would be needed in the visible and near-infrared to explain any (controversial) out-of-phase variability in these bands with respect to solar proxies such as the DSA (Harder et al. 2009) because our model relating the SSI to sunspot-area change does not produce lags in the SSI of more than a few months (as mentioned in Section 4.1). Thus, such out-of-phase variability would instead be incorporated into the residual contribution and not the impulse response itself.

Let us stress again that in a non-parametric model, the residual contribution cannot be neglected, as it was in previous studies, since this affects the determination of the impulse response itself, as shown in Section 3.2. The example shown in Figure 4 clearly indicates the need for using Equation (4). If, however, we ignore the residual contribution then reasonably good reconstructions can still be achieved by forcing Equation (1) to the data, as in Preminger & Walton (2007). However, the quality of the reconstructed output in this case does not necessarily reflect the quality of the model parameters, with the appearance of long and unphysical tails in $h(t)$.

With a nonzero residual contribution such as what we find, the SSI cannot be described solely by the sunspot areas because we have no model for describing how the residual contribution itself varies. Since numerous attempts have been made to infer past values of the TSI or SSI from sunspot data on a daily to monthly basis (e.g., Solanki & Fligge 1999; de Toma et al. 2001; Pap et al. 2002; Balmaceda et al. 2007; Preminger & Walton 2007), it is insightful to determine what spectral bands require a significant residual contribution. To quantify this, we compute the ratio between the amplitudes of $r(t)$ and of $y(t)$ after decimating the latter to the 27-day sampling rate of $r(t)$. The results are summarized in Figure 8, which shows that the residual contribution represents at least 50% of the long-term variation in the UV bands and approximately 100% of the long-term variation in the visible and NIR.

The surprisingly large magnitude of the residual contribution may have several causes. As mentioned previously, this contribution is required to describe physical out-of-phase variations of the SSI on long timescales, namely those of several months and longer.

Out-of-phase variations on shorter timescales have been well documented (e.g., Vigouroux et al. 1997; Preminger & Walton 2005) and are captured by the impulse response itself. Longer-term variations may be caused by ephemeral regions,

whose contributions vary cyclically with the solar cycle but start before and end after the corresponding cycle (Harvey 1992). Geometrical effects may also contribute: since the average latitudinal position of sunspots decreases during each solar cycle, the response of the SSI to sunspot variations slightly differs between the rising and decaying phase of the solar cycle. This essentially means that the impulse response is not completely time-invariant.

A different cause may be the presence of uncorrected instrumental drifts in the SSI, which the model cannot fit and thus defers to the residual contribution. Such drifts are more likely to occur in the visible and near-infrared bands, where the relative solar-cycle variability is lower than that in the more energetic shorter wavelengths. However, since we find the residual is comparable in the visible band and the TSI, which share similar properties but have different long-term uncertainties (with the TSI measurements being considerably more accurate and stable on solar-cycle timescales), we conclude that instrumental effects are likely not a major contributor to the residuals shown.

Finally, since the residual contribution captures what the linear impulse response model cannot fit, what we observe may be the consequence of the limitations of this linear model with respect to a Sun that is not fully linear, time-invariant, and causal. These possible effects are discussed in more detail in Section 4.5.

We conclude here that the residual contribution, which represents an important part of the variability, probably captures both instrumental effects and physical processes that cannot be described by the impulse response. Therefore it is risky (if not impossible) to draw conclusions about long-term (i.e., solar cycle) variability of the SSI and TSI simply by considering the linear and time-invariant short-term response to changes in sunspot area, because a significant portion of the solar variability, particularly in some wavelength ranges, is only partly related to these changes.

4.3. Energy Balance Due to Sunspot-area Variations

In principle, the temporal integration of the impulse response in Figure 5 could enable energy-budget estimates and determine whether the average net contributions from facular brightenings exceed those from sunspot dimmings over the duration of the response. This is clearly the case for the shorter (EUV, FUV, and MUV) wavelength bands in Figure 5 and for the solar indices other than the TSI in Figure 6, as these impulse responses are always positive. At longer wavelengths and for the TSI, the net contribution to brightness can be negative when integrated over the few solar rotations where the magnitude of $h(t)$ exceeds the noise limit, suggestive of a net energy loss due to the emergence of new sunspot-area flux. Temporal integrations of the responses for positive and negative values of $h(t)$ are shown as a function of wavelength in Figure 9 along with the sum of the two.

As the noise limit prohibits lengthier integrations, drawing conclusions on the net radiative energy balance due to increases in sunspot area remains uncertain. Additionally, extending these short-term energy-balance results to solar-cycle timescales, as suggested by Woods et al. (2015), is even less certain since many solar-cycle effects are absorbed into the relatively large residual component $r(t)$ rather than the impulse response itself. We thus conclude that it is particularly difficult to draw solar-cycle energy-balance conclusions via this linear impulse-response
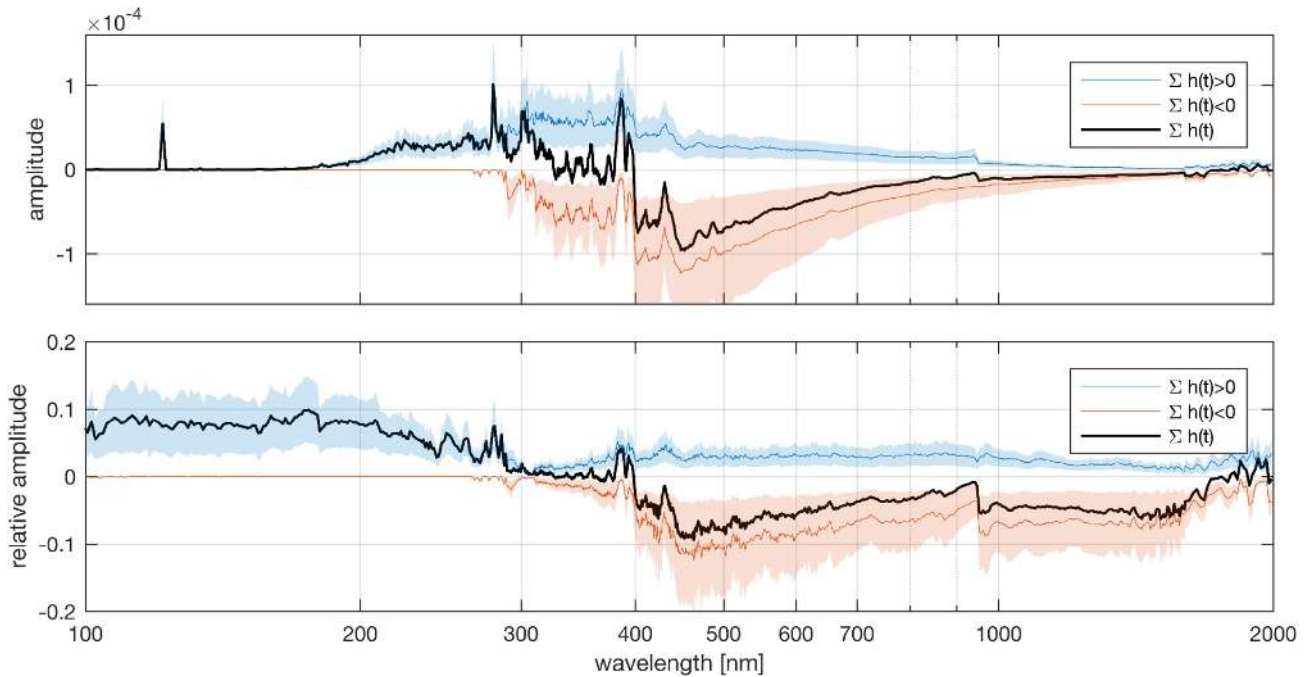
Figure 9. Amount of spectral solar irradiance change per unit increase in the DSA. The blue (resp. red) curves indicate the amount of irradiance change that is caused by an excess (resp. deficit) of irradiance. The black curve is the sum of the two and expresses the total irradiance budget as a function of wavelength. The upper plot is in absolute units (W nm^{-1} m^{-2} per 100 μhem), while in the bottom plot all quantities have been normalized with respect to the solar-cycle variability of the SSI. Uncertainties of $\pm$ one standard deviation are represented by shaded areas. For greater clarity, uncertainties are not indicated for the net irradiance budget (black); however, they are of the same order of magnitude as the ones given for the excess and deficit. Some of the discontinuities, such as the dip near 930 nm, are instrumental, for instance coinciding with transitions between different detectors.

approach. While the approach can determine short-term (few solar rotations) net energy responses to flux emergence for different wavelength bands, we find that the method is unable to determine in-phase or out-of-phase responses for the SSI on solar-cycle timescales.

4.4. Significance of Precursors

Our transfer function is causal in the sense that it assumes the SSI cannot start to vary before a sunspot appears. However, small precursors may exist, for example, with ephemeral regions that occur before the sunspot emerges (Martin & Harvey 1979). The impulse response obtained by Preminger & Walton (2005) actually shows variations in $h(t)$ for negative times. However, since their impulse response includes solar-rotational variations as well as contamination from (unfitted) residuals, an ambiguity in start time may explain this effect: in their model, $t = 0$ corresponds to the occurrence of the maximum increase in sunspot area, which is influenced by sunspot position relative to the center of the disk (when its area is largest). Thus, increases in the SSI may be observed at least a few days earlier (i.e., for $t < 0$) when the sunspot becomes visible on the east limb. Such ambiguities are greatly reduced in our approach of using 27-day snapshots to avoid such potential solar-rotational effects.

To check for the presence of precursors we estimated the impulse response after artificially delaying the solar response by an integer number of solar rotations, typically one to six. Although the impulse response fluctuates before the actual emergence of the sunspot, we never found these variations to be significant at the 0.05 level. We thus conclude that there is no evidence for precursors in the records we utilize here.

4.5. Limitations—Nonlinearity in the Impulse Response

A major assumption behind our analysis is the linearity of the solar response to sunspot-area variations. While several studies suggest that this response should be mostly linear (Foukal & Lean 1990; Lean et al. 1995; Preminger & Walton 2005), nonlinear effects are expected to become significant for high levels of solar activity (Brown & Evans 1980; Foukal 1998; Solanki & Fligge 1999).

From physical considerations, the linear proportionality between DSA and SSI is expected to level off at high solar activity because the emergence of new sunspots is increasingly hindered by the presence of nearby ones. To check for such a nonlinear signature in the impulse response, we estimated the latter separately for three regimes of the solar cycle: weak activity (81-day average of the DSA < 250 μhem), medium activity ($250 <$ DSA < 750 μhem), and high activity (DSA > 750 μhem). The uncertainties on the results are larger than before because the sample sizes are smaller and the samples are not continuous in time.

Figure 10 summarizes the impact of the average level of solar activity by showing the impulse response at two solar-rotation times, namely at $h(t = 0)$ and at $h(t = 27$ days). For both of these times, we plot the impulse response for the FUV, the visible, and the TSI for each of the three different net sunspot-area values, which are indicative of different levels of solar activity through the solar cycle. This plot reveals a nonlinear response for different phases of the solar cycle.

For the UV bands, we find that the initial $t = 0$ value of the impulse response, which is positive due to net solar brightening at these wavelengths, decreases by up to 40% for the larger net sunspot areas typical of solar maximum. In the visible bands and for the TSI, where this initial response is negative due to
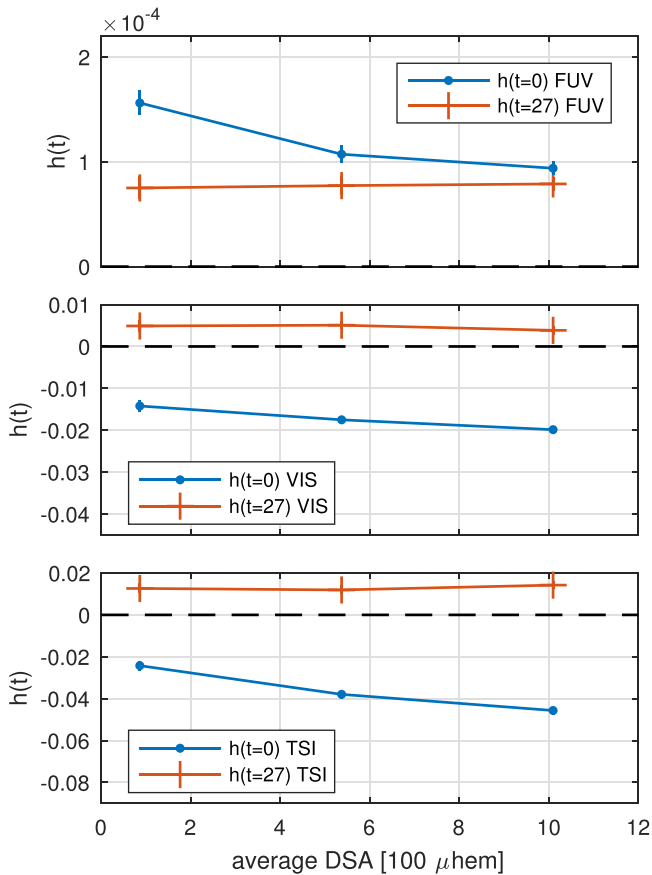
Figure 10. Magnitude of the impulse response $h(t)$ at $t = 0$ and one solar rotation later for the FUV band (upper plot), the visible band (middle plot), and the TSI (lower plot). Three values are shown, corresponding to three levels of solar activity. The error bars represent $\pm 1\sigma$.

overall solar darkening, the impulse response becomes stronger (that is, more negative) toward solar maximum. This means that an equal increase in sunspot area produces a stronger darkening at solar maximum than at solar minimum. Such a behavior of the impulse response function obtained for different activity levels indicates that the ratio of facular area to sunspot area is lower during periods of higher solar-activity levels. This is in line with previous observations; see, for example, Figure 2 in (Foukal 1993) and Figure 1 in (Foukal 1998).

Interestingly, after one or more solar rotations the impulse response is independent of the level of activity. This is because while the initial response is caused by an increase in sunspot area and any associated faculae, the later response is mostly caused by the decay products of the emerged sunspots, and those products do not depend on the level of solar activity.

Figure 10 shows that the nonlinear response at $t = 0$ is not just a matter of applying a static nonlinear correction to the input (e.g., using the square root of the DSA as input). If this had been the case, then the magnitude only and not the shape of the impulse response $h(t)$ would have depended on the level of solar activity.

Figure 10 indicates that the dynamics itself (i.e., the characteristic response time) is solar-cycle dependent. The properties of such nonlinear systems cannot be properly inferred from the usual scatterplots between the driver and the solar response; they require a system identification approach with nonlinear models

(e.g., Nelles 2001). Unfortunately, such techniques are much more demanding in terms of data quality and volume; whether they can be meaningfully applied to the 12 years of data used here remains to be investigated.

Nonlinear effects may also manifest themselves in other aspects, such as a solar-cycle dependence of the characteristic response time, i.e., an absence of time-invariance. Currently, with the relatively limited amount of available observations and the omnipresence of noise, it is unlikely that such second-order effects can be properly quantified. In this paper, while we acknowledge the possibility of these nonlinear effects, we thus consider only the average linear impulse response and its deviations.

5. Conclusions

The primary motivation of this study was to use the powerful framework of linear system theory to better understand the response of the SSI to sunspot-area variations.

We have discovered and presented several significant methodological improvements to the linear impulse-response model applied to sunspot-area changes (flux emergence) and the corresponding effects on the SSI. We have obtained a compact representation of the impulse response of the SSI to sunspot area that is independent of artifacts due to solar rotation via analyzing solar-rotational "snapshots." We have shown that a non-parametric model of the response of SSI cannot be entirely described by $h(t)$, but that a residual term is required. This important term, absent in prior such works, describes a major fraction of long-term variability. The existence of this residual term implies that one cannot describe the full variability of the SSI by solely using the sunspot area as input. This has implications on the use of empirical models that rely purely on the sunspot number: while such models perform well for short-term variability (days to months), they may fail to properly describe long-term (solar-cycle) variability.

Additionally, we have identified several physical results regarding the solar response to sunspot-area increases. The impulse response presented quantifies the global brightening observed in the UV. In the visible and near-infrared bands, an initial deficit of radiation is followed by a brightening, which can be observed for three to five solar rotations. Uncertainties associated with the relatively short (12 year) SSI data set time span prevent observation of potentially lengthier decay responses.

Nevertheless, we estimate a net energy budget for short-term variability as a function of wavelength, showing in which spectral bands sunspot-darkening outweighs facular brightening. This energy budget includes uncorrected instrumental effects inherent in the data, which may limit its interpretation.

We find that the response of the SSI is nonlinear with the level of solar activity. This effect is mostly observed during the initial occurrence of the spot; the facular brightening that follows at solar-rotation times $t > 0$ is independent of the level of activity. This indicates that the dynamical response of the Sun is nonlinear and cannot be corrected simply by applying a rescaling to the sunspot area.

There are two fundamental limitations to the conclusions presented here. *Data-length limitations* prevent us from drawing conclusions regarding solar-cycle variability since we have only relatively short durations of measurements of SSI variations. *Methodological limitations* of our linear model, which captures most of the short-timescale dynamics of the

Sun, prevent it from reproducing variations equally well during solar maximum and solar minimum. A nonlinear impulse-response model is required since the response of the SSI to sunspot area is not instantaneous, and therefore requires a convolutive model. This will be the subject of future endeavors based on the results presented here.

Facility: SORCE; ADS.

Software: Matlab.

ORCID iDs

T. Dudok de Wit ⓘ https://orcid.org/0000-0002-4401-0943

References

Antsaklis, P. J., & Michel, A. N. 2005, Linear Systems (Boston: Birkhäuser)
Ball, W. T., Haigh, J. D., Rozanov, E. V., et al. 2016, NatGe, 9, 206
Balmaceda, L., Krivova, N. A., & Solanki, S. K. 2007, AdSpR, 40, 986
Brown, G. M., & Evans, D. R. 1980, SoPh, 66, 233
Coddington, O., Lean, J. L., Pilewskie, P., Snow, M., & Lindholm, D. 2016, BAMS, 97, 1265
de Toma, G., White, O. R., Chapman, G. A., et al. 2001, ApJL, 549, L131
Domingo, V., Ermolli, I., Fox, P., et al. 2009, SSRv, 145, 337
Dudok de Wit, T. 2011, A&A, 533, A29
Dudok de Wit, T., Bruinsma, S., & Shibasaki, K. 2014, JSWSC, 4, A06
Dudok de Wit, T., Kopp, G., Fröhlich, C., & Schöll, M. 2017, GeoRL, 44, 1196
Eddy, J. A. 1976, Sci, 192, 1189
Ermolli, I., Matthes, K., Dudok de Wit, T., et al. 2013, ACP, 13, 3945
Foukal, P. 1993, SoPh, 148, 219
Foukal, P. 1998, ApJ, 500, 958
Foukal, P. 2013, Solar Astrophysics (3rd ed.; Weinheim: Wiley-VCH)
Foukal, P., & Lean, J. 1990, Sci, 247, 556
Fröhlich, C., & Lean, J. 2004, A&ARv, 12, 273
Harder, J. W., Fontenla, J. M., Pilewskie, P., Richard, E. C., & Woods, T. N. 2009, GeoRL, 33, L07801
Harvey, K. L. 1992, in ASP Conf. Ser. 27, The Solar Cycle, ed. K. L. Harvey (San Francisco, CA: ASP), 335
Hickey, J. R., Stowe, L. L., Jacobowitz, H., et al. 1980, Sci, 208, 281
Kopp, G. 2016, JSWSC, 6, A30
Lean, J. L. 2000, GeoRL, 27, 2425
Lean, J. L., Beer, J., & Bradley, R. 1995, GeoRL, 22, 3195
Ljung, L. 1997, System Identification: Theory for the User (2nd ed.; Englewood Cliffs, NJ: Prentice-Hall)
Martin, S. F., & Harvey, K. H. 1979, SoPh, 64, 93
Nelles, O. 2001, Nonlinear System Identification (Berlin: Springer)
Oberländer, S., Langematz, U., Matthes, K., et al. 2012, GeoRL, 39, 1801
Pap, J. M., Turmon, M., Floyd, L., Fröhlich, C., & Wehrli, C. 2002, AdSpR, 29, 1923
Preminger, D. G., Chapman, G. A., & Cookson, A. M. 2011, ApJL, 739, L45
Preminger, D. G., & Walton, S. R. 2005, GeoRL, 32, 14109
Preminger, D. G., & Walton, S. R. 2006a, GeoRL, 33, L23108
Preminger, D. G., & Walton, S. R. 2006b, SoPh, 235, 387
Preminger, D. G., & Walton, S. R. 2007, SoPh, 240, 17
Rottman, G. 2005, SoPh, 230, 7
Schoukens, J., Vandersteen, G., Barbé, K., & Pintelon, R. 2009, Eur J Control, 15, 260
Schreiber, T., & Schmitz, A. 2000, PhyD, 142, 346
Seleznyov, A. D., Solanki, S. K., & Krivova, N. A. 2011, A&A, 532, A108
Shapiro, A. I., Solanki, S. K., Krivova, N. A., et al. 2017, NatAs, 1, 612
Shapiro, A. I., Solanki, S. K., Krivova, N. A., Tagirov, R. V., & Schmutz, W. K. 2015, A&A, 581, A116
Shapiro, A. I., Solanki, S. K., Krivova, N. A., Yeo, K. L., & Schmutz, W. K. 2016, A&A, 589, A46
Solanki, S. K., & Fligge, M. 1999, GeoRL, 26, 2465
Solanki, S. K., Krivova, N. A., & Haigh, J. D. 2013, ARA&A, 51, 311
Spruit, H. 2000, SSRv, 94, 113
Vigouroux, A., Pap, J. M., & Delache, P. 1997, SoPh, 176, 1
Wehrli, C., Schmutz, W., & Shapiro, A. I. 2013, A&A, 556, L3
Woods, T. N., Eparvier, F. G., Bailey, S. M., et al. 2005, JGR, 110, A01312
Woods, T. N., Snow, M., Harder, J., Chapman, G., & Cookson, A. 2015, SoPh, 290, 2649
Yeo, K. L., Krivova, N. A., Solanki, S. K., & Glassmeier, K. H. 2014, A&A, 570, A85
Yeo, K. L., Solanki, S. K., Norris, C. M., et al. 2017, PhRvL, 119, 091102