Data Fusion of Total Solar Irradiance Composite Time Series Using 41 years of Satellite Measurements

J.-P. Montillet[1], W. Finsterle[1], G. Kermarrec[2], R. Sikonja[3], M. Haberreiter[1],W. Schmutz[1], T. Dudok de Wit[4]

[1]Physikalisch-Meteorologisches Observatorium Davos/World Radiation Center (PMOD/WRC), Davos, Switzerland
[2]Geodätisches Institut, Leibniz Universität Hannover, Hannover, Germany
[3]Department of Computer Science, Eidgenössische Technische Hochschule (ETH), Zurich, Switzerland
[4]Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, CNRS, CNES and University of Orléans, Orléans, France

Key Points:

- To provide a new composite TSI time series (PMOD Data Fusion v1) with 41 years of satellite observations relying on data fusion
- A comprehensive time-frequency analysis to characterise the solar cycle and the stochastic noise
- Full investigation of variations at solar minima to distinguish between stochastic noise and possible underlying phenomena linked to the solar activity

Corresponding author: J.-P. Montillet, jean-philippe.montillet@pmodwrc.ch

Abstract
Since the late 70's, successive satellite missions have been monitoring the sun's activity, recording the total solar irradiance (TSI). Some of these measurements last for more than a decade. It is then mandatory to merge them to obtain a seamless record whose duration exceeds that of the individual instruments. Climate models can be better validated using such long TSI records which can also help provide stronger constraints on past climate reconstructions (e.g.,back to the Maunder minimum). We propose a 3-step method based on data fusion, including a stochastic noise model to take into account short and long-term correlations. Compared with previous products, the difference in terms of mean value over the whole time series and at the various solar minima are below 0.2 W/m^2. Next, we model the frequency spectrum of this 41-year TSI composite time series with a Generalized Gauss-Markov model to help describing an observed flattening at high frequencies. It allows us to fit a linear trend into these TSI time series by joint inversion with the stochastic noise model via a maximum-likelihood estimator. Our results show that the amplitude of such trend is $\sim -0.009 \pm 0.010$ $W/(m^2 yr)$ for the period 1980-2021. These results are compared with the difference of irradiance values estimated from two consecutive solar minima. We conclude that the trend in these composite time series is mostly an artefact due to the coloured noise.

1 Introduction

Monitoring the Earth's radiation budget is key to understand the anthropogenic contribution to climate forcing (Kren, 2015). Total solar irradiance is Earth's dominant energy input. Global temperature and TSI are linked by the energy equilibrium equation for the Earth system. As summarized by Schmutz (2021), the derivation of this equation with respect to a variation of the solar irradiance has two terms: a direct forcing term, which can be derived analytically and quantified accurately from the Stefan-Boltzmann law, and a second term, describing indirect influences on the surface temperature. If a small TSI variation should force a large temperature variation, then it has to be the second indirect term that strongly amplifies the effect of the direct forcing. This amplification mechanism has been debated in the scientific community for the past two decades (Rind et al., 2014; Shapiro et al., 2017; Egorova et al., 2018; Schmutz, 2021), because it will most likely call for a strong modification of the models that describe the Earth's climate response to variations in the solar radiative output. On shorter time scales (e.g., weekly), the existence of trend in the measurements could on a longer timescale (e.g., yearly) bias significantly the analysis of a solar phenomena (e.g., estimation of a new solar minima). Therefore, it is important to produce robust and reliable TSI composite time series using all the observations available recorded by successive space instruments spanning 4 decades. The satellite measurements show that the TSI varies at all timescales with a pronounced signature of quasi-periodicity of approximately 11 years (Fröhlich et al., 1997; Kopp, 2016). Timescale variations can be classified in subdaily (minutes to hour), daily to weekly, and yearly to one solar cycle. Major mechanisms, such as the evolution of magnetic features on the solar surface, dominating each timescale are complex and still under investigation within the solar physics community (Yeo et al., 2017; Xiang, 2019). Several studies (Fontenla et al., 1999; Kopp & Lean, 2011; Yeo et al., 2021) have shown that TSI variations on timescales of hours are a combination of the sunspots blocking and the intensification due to bright faculae, plages and other elements. This contributes to the difficulty of forecasting and modelling the solar cycle. All satellite observations are limited in time, making composites the key to investigation over several decades. Merging all these observations is a difficult exercise with both a scientific and a statistical challenge (Dudok de Wit et al., 2017). Previous approaches (Willson, 1997; Fröhlich & Lean, 2004; Mekaoui & Dewitte, 2008) produced TSI composite time series by daisy chaining all the available TSI observations without including any models of the stochastic noise properties. The first methodology which relied on some knowledge of the underlining noise

characteristics was developed by Dudok de Wit et al. (2017), including a data-driven noise model and a multiscale decomposition.

We are here presenting a 3-step method. The first step relies on data fusion of multiple observations based on a Bayesian framework and Gaussian processes. Our composite spanning the last 4 decades is obtained in the second step by daisy chaining the sub-time series resulting from the first step. The last step is the application of a wavelet filtering to correct some unwanted correlations in the fused observations (i.e. bandwidth noise). The robustness of our approach is guaranteed via a careful modelling of the TSI observations during the data fusion process. Various assumptions formulated by data scientists, can introduce biases in the data analysis. Some algorithms (Willson, 1997; Fröhlich & Lean, 2004; Mekaoui & Dewitte, 2008) based on daisy chaining the raw TSI observations required the choice of the most trustworthy instrument, hence introducing a bias toward preconceived ideas of how the TSI should vary. Note that our data fusion process is merging dataset from subsequent solar missions only based on a few stochastic noise assumptions. It circumvents the weakness of choosing the most trustworthy instrument when performing the daisy chain on the TSI observations from various instruments, which could influence towards preconceived ideas of how the TSI should vary (Dudok de Wit et al., 2017).

Finally, recent studies (Scafetta et al., 2020; Dudok de Wit & Kopp, 2020; Schmutz, 2021) have debated about the existence of a trend in the composite time series. If it exists, the origin of this trend in the TSI observations is unknown: one could speculate that it could be caused by a drift in the solar cycle peaks' amplitude while the minima could all remain at the same level. Another possibility is the presence of an unknown diffusion process which could generate a transient signal making variations in the solar minima. Dudok de Wit and Kopp (2020) argue in favor of an artifact generated by unwanted noise looking at the difference between consecutive solar minima. Here, we go further by performing a time-frequency analysis of various TSI composite time series produced with various techniques, including our new product. We focus on describing the stochastic noise properties within these 40-year long time series. We use this knowledge to model the TSI composites and to conclude on the existence of a long-term trend.

2 Description of The Raw Dataset

Table 1 displays the instruments and the processing centers providing the observations relative to the various missions used in this study. The data processing, including corrections for all a priori known influences such as distance from the sun (normalized to 1 AU), radial velocity to the sun, and thermal, optical, and electrical corrections, are usually implemented by each processing center, leading to level-1 time series. Most of these instruments observe on a daily basis, with occasional interruptions and outliers. Usually, one to three of them are operating simultaneously, although some days are devoid of observations. Note that *PMODv21a* is the new VIRGO/SOHO dataset released in March 2021 by PMOD using a new software described in Finsterle et al. (2021). *PREMOS (v1)* is the released version described in Schmutz et al. (2013). ERBE and HF dataset are retrieved from the PMOD archive and the corrections made by C. Fröhlich, which are explained in Fröhlich (2006).

Figure 1 displays the observations from each mission spanning a specific period of time. All the space missions have provided TSI observations with a different sampling rate. Recent instruments make several observations per day (with a cadence of up to 50 s for TIM). Earlier radiometers such as ERBE observes the sun once every 14 days for 3 min on average, so that the stochastic noise properties of such sensors are different from higher recording rate instrument. Note that *active* in Table 1 means that the instrument is still operating. The dataset for these missions ends up in March 2021 for this study.

Mission/Experiment/Instrument	Version	Start Date	End Date
HF/NIMBUS-7 ERB	-	11/1978	1/1993
ERBE/ERBS	-	10/1984	8/2003
VIRGO/SOHO	PMODv21a	01/1996	active
PREMOS/PICARD	v1	06/2010	03/2014
ACRIM1/SMM	1	2/1980	7/1989
ACRIM2/UARS	7/14	10/1991	9/2000
ACRIM3/ACRIMSAT	11/13	04/2000	11/2013
TIM/SORCE	18	02/2003	02/2020
TIM/TSIS	3	11/01/2018	active

Table 1: Overview of the datasets used in this study including the starting & ending dates for each mission and the latest version released by the various centers.

3 The 3-step Method to Produce the 41-year Long TSI Composite

3.1 Step 1: Merging Multiple Dataset with Data Fusion

Data fusion is the process of integrating multiple data sources to produce more consistent, accurate, and useful information than that provided by each individual data source alone. The process has found many applications in various areas ranging from industry to geosciences and solar science (Cocchi, 2019).

Let us call the observations to merge $(a(t_i), b(t_i), c(t_i))$ (with $\{i = [1, n]\}$) recorded from 3 different instruments. The noise for each observation is additive and uncorrelated between the instruments. The model of the observations is defined such as:

$$\begin{cases} a(t_i) = s(t_i) + \epsilon_a(t_i), & \epsilon_a(t_i) \sim \mathcal{N}(0, \sigma_a^2) \\ b(t_i) = s(t_i) + \epsilon_b(t_i), & \epsilon_b(t_i) \sim \mathcal{N}(0, \sigma_b^2) \\ c(t_i) = s(t_i) + \epsilon_c(t_i), & \epsilon_c(t_i) \sim \mathcal{N}(0, \sigma_c^2) \end{cases} \tag{1}$$

where ϵ_a, ϵ_b and ϵ_c are zero-mean Gaussian distributed random variables (with variance σ_a^2, σ_b^2 and σ_c^2 respectively) modelling the noise properties intrinsic to each instrument. The data fusion algorithm aims at merging the observations available at each epoch t_i in order to get a reliable estimate of the true signal s, i.e. the solar activity (Feynman, 1982). We formulate the following assumptions: i) the solar cycle is an unknown process (i.e. not a perfect sinusoidal signal with 11.5 year cycle) and its variations are random (no a priori knowledge). Physically, it means that two or more radiometers monitor the solar activity from a different distance due to different orbits, but monitoring the same underlying information on the solar cycle. The model of s is a Gaussian process (GP) with zero mean and a covariance function k_θ (or kernel). A GP can be generally defined as a finite sum of random variables normally distributed where the overall distribution is a multivariate normal distribution (Kolar et al., 2020); ii) we consider

the noise on the measurements zero-mean Gaussian distributed, we can then estimate the parameters of the model of $s(t)$ via maximum likelihood estimator (MLE). Therefore, we have $\mathbf{s} \sim GP(0, k_\theta(t_i, t_i)_{\{i=[1,n]\}})$, with n the number of samples in the various measurements a, b and c. The parameters of $s(t)$, expressed in θ, are selected by maximizing the log-likelihood $\log p(\mathbf{y}|\mathbf{x})$, where $\mathbf{x}$ and $\mathbf{y}$ are the concatenation of times, i.e. $\mathbf{x} = [t_i, t_i]$, and corresponding corrected observations, i.e. $\mathbf{y} = [a(t_i), b(t_i), c(t_i)]$. The main limitation of GPs is that given n observations, the inverse of the n-by-n covariance matrix must be computed. Time complexity of such operation is of the order of $O(n^3)$, which is computationally expensive for long records. Some of these missions have been recording data over two decades generating large data sets. To overcome this limitation, we approximate the exact GPs by utilizing sparse Gaussian processes (SGP), yielding a maximization problem of the lower bound of $\log p(\mathbf{y}|\mathbf{x})$ following Bauer et al. (2016):

$$\log p(\mathbf{y}|\mathbf{x}) \geq -\frac{1}{2}\mathbf{y}^T(\mathbf{Q}_\theta+\sigma^2\mathbf{I})^{-1}\mathbf{y}-\frac{1}{2}\log|\mathbf{Q}_\theta+\sigma^2\mathbf{I}|-\frac{n}{2}\log(2\pi)-\frac{1}{2\sigma^2}tr(k_\theta(\mathbf{x},\mathbf{x})-\mathbf{Q}_\theta) \quad (2)$$

where tr is the trace operator, $Q_\theta = k_\theta(\mathbf{x},\mathbf{u})k_\theta(\mathbf{u},\mathbf{u})^{-1}k_\theta(\mathbf{u},\mathbf{x})$, $\mathbf{u}$ is a vector of inducing points to learn about the stochastic properties of the data, which allows to take into account long-term and short-term correlations in the observations and are a reasonable approximation of s. This subset of observations is used to estimate the initial parameters in θ. $\mathbf{I}$ is the identity matrix, with $\sigma^2\mathbf{I}$ the noise component of the covariance matrix (assuming uncorrelated measurements) formulated as $diag([\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_a^2, \sigma_b^2, \sigma_c^2, ...])$. Next, we estimate the kernel k_θ by maximizing the right-hand-side of Eq. (2) with respect to $\mathbf{u}$ and θ. Further mathematical simplifications to estimate the kernel are voluntarily left out for clarity, but readers can refer to Kolar et al. (2020) for additional information. We emphasize that the number of inducing points defines the size of the matrix $\mathbf{Q}_\theta$ which must be inverted in the maximization of Eq. (2). A large number of points is necessary to avoid smoothing completely the short-term and long-term correlations due to the difference in the recording rate of the instruments. The computational complexity is in the order of $O(nm^2)$ (with m the number of inducing points, $m \ll n$). Therefore, we are limited by the computing resources available when dealing with a large matrix (i.e. over $m = 3000$). Now, the number of inducing points varies due to the size of the input datasets (i.e. size of the boxes defined in Figure 1). In order to see the influence of this parameter on the quality of the fused time series, Figure Appendix C.1 shows the variations in both time and frequency when fusing VIRGO/SOHO, TIM/SORCE and ACRIM3/ACRIMSAT (box 8 in Figure 1). Figure Appendix C.2 displays the associated power spectrum density (PSD). It is difficult to find an optimal number, because above 1500 points the time fluctuations do not show much differences, but the PSD is still varying (i.e. continuity of the spectrum, amplitude of the frequencies associated with the solar cycle). We chose 2000 points, which provide a good balance between computational time and accuracy. Now, most of the sub-time series have a length greater than 7 years, therefore one can select 2500 points or more if necessary. The shortest time series is when fusing PREMOS/PICARD, VIRGO/SOHO and TIM/SORCE (box 7). We have $\sim$ 1400 observations. In this case, we use 1300 points. Note that the number of inducing points for the fusion of TIM/TSIS and VIRGO/SOHO is also constrained. This is due to the availability of both products at the time of writing this article.

As described in Section 2, each instrument records the data with different sampling rate. The fusion requires regularly sampled records with no gaps. We first regrid all the datasets with a sampling rate higher than 1 day. The datasets recorded with a lower rate are (linearly) interpolated. Note that the starting date of the composite time series is defined by the fusion between HF and ACRIM 1 which is 2/1980.

Nonetheless, events resulting from short-term variations in solar activity lasting less than a few days (e.g. solar flares) are relatively difficult to fuse. The radiometers on board of the various missions at time t_i may not have recorded exactly the same event due to different distances (i.e. different orbits) and also because of differences in the observation time (i.e. sampling rate). The fusion of these short-term solar variations results gen-

erally in keeping only major events or underneath long-term solar events recorded by all the instruments at t_i.

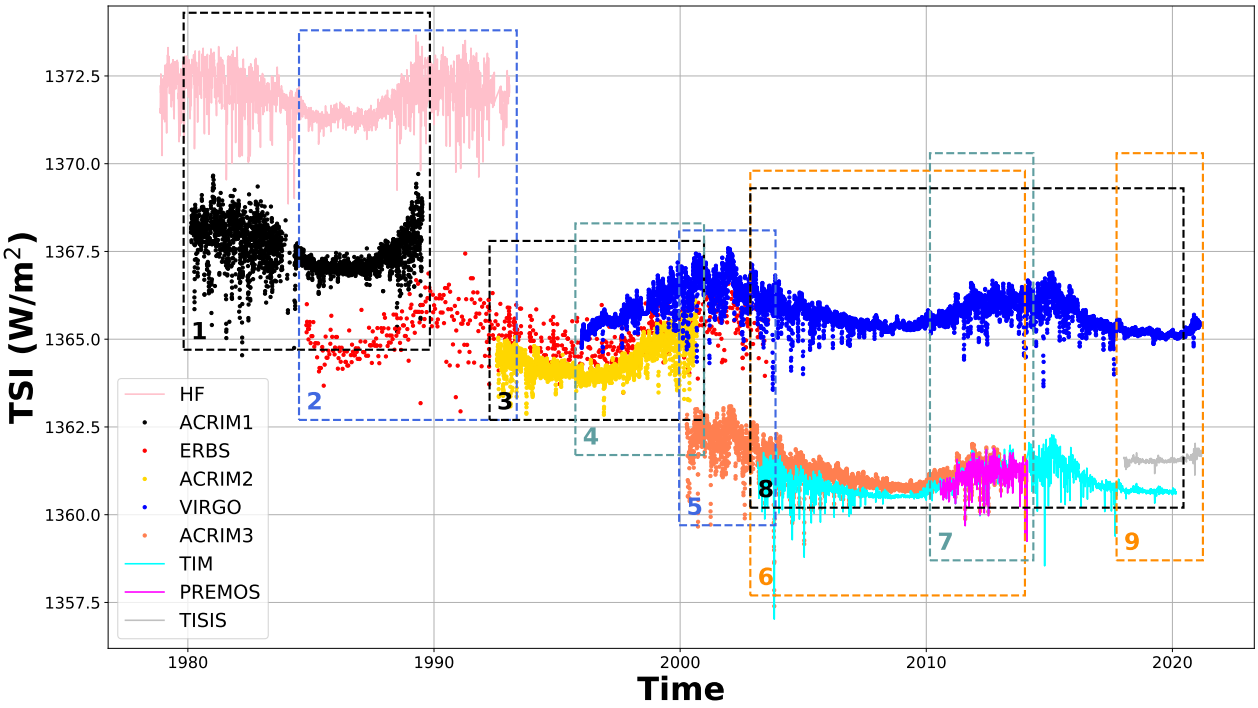

Figure 1: The various satellite missions recording TSI observations since 1980. We perform the fusion using the observations included in each box (dash lines).

Last, we should distinguish in the following between stochastic and solar noises in order to avoid any confusion for the reader. We refer as stochastic noise, the statistical definition of random processes which includes short and long-term correlations (i.e. white and coloured noise). Solar noise results from the photospheric activity associated with granules varying at different timescales over a few hours (e.g., sunspots) to a decade (e.g., solar cycle) generating fluctuations in the recorded irradiance values.

3.2 Step 2: Producing the 41-year composite time series with a modified adaptive filter

To perform the data fusion, we first select all the periods where at least two missions overlap for more than 6 months (see boxes in Figure 1). With a shorter overlapping time, simulations have shown that the fusion is not optimal due to the limited number of inducing points. For each overlapping period, we fuse the time series corresponding to different missions/instruments together in order to obtain the sub-time series.

We produce q partially overlapping composite time series (y_q) with associated uncertainties (α_q^2). We use a modified adaptive algorithm (Haykin, 2004) to daisy-chain all the sub-time series and build the 41-year composite such as:

$$\begin{cases} y(t_i) = & y_1(t_i) * w(t_i) + y_2(t_i) * (1 - w(t_i)) \\ w(t_i) = & \alpha_1^2(t_i)/(\alpha_2^2(t_i) + \alpha_1^2(t_i)) \\ \bar{\alpha_1^2} \leq \bar{\alpha_2^2} & \\ \alpha^2(t_i) = & 0.5 * (\alpha_2^2(t_i) + \alpha_1^2(t_i)) \end{cases} \tag{3}$$

with t_i the time spanning the period 1978-2021, with the sampling of 1 day. The two time series overlapping are y_1 and y_2 and associated uncertainties α_1^2 and α_2^2 respectively. $\bar{\alpha_1^2}$, $\bar{\alpha_2^2}$ are the average of the uncertainties over the overlapping time for y_1 and y_2. Note that y_1 is chosen in order to satisfy the condition $\bar{\alpha_1^2} \leq \bar{\alpha_2^2}$. We define w (in $[0, 1]$) using the denominator $(\alpha_2^2(t_i)+\alpha_1^2(t_i))$ in order to avoid a divergence of w. We exclude the case for which $\alpha_1^2(t_i) = \alpha_2^2(t_i) = 0$.

Now, the mean value of each sub-time series resulting from the data fusion process is relative to the lowest mean value of the input TSI datasets. We end up with a different mean value for each sub-time series. Before applying the modified adaptive algorithm on two consecutive time series, we scale the second sub-time series using the common period between the two time series. It results in a TSI composite time series with an arbitrary mean value. To obtain the correctly scaled TSI composite, we employ the TSI value by Prša et al. (2016). The nominal TSI in Prša et al. (2016) was derived as the averaged TSI value of Solar Cycle 23. This approach is also applied here, i.e. we determine the average TSI for Solar Cycle 23 of the new composite and scale it to the nominal TSI value. As such the new TSI composite is consistent with the nominal TSI of 1361 W/m^2 as recommended by the IAU 2015 Resolution B3. Finally, the last step includes a wavelet filter in order to smooth the correlations introduced by the data fusion. The effect of these correlations in time and frequency domains is discussed in the next section.

3.3 Step 3: Filtering the Composite with a Wavelet Filter

The data fusion process acts as a low pass filter, which results in a modification of the PSD of the composite time series at high frequencies. Possible reasons are to be associated with the number of inducing points or other parameters as discussed in Section 3.1. This drawback is unwanted: 1) it may mask some peaks at high frequencies linked with daily components; 2) the solar minimum can be affected in the time domain by the presence of long-term correlations. Unfortunately, this effect is theoretically unpredictable. We propose to reconstruct empirically the high frequencies of our fused time series. To that aim, the wavelet variance (WV) provides a robust mathematical framework to perform the rescaling of the filtered noise (Abry & Veitch, 1998). More specifically, we decompose the time series into an ensemble of records whose spectral content is concentrated in a specific frequency band. By adding them together, the original time series is recovered. We make use of the Maximum Overlap Discrete Wavelet Transform (MODWT), which has some advantages over the usual discrete wavelet transform: it permits to avoid a downsampling process, unfavourable for an analysis at high scales (Percival & Guttorp, 1994).

In the input parameters, we choose the least asymmetric wavelet ($LA(4)$) with 8 scales which provides coefficients that are approximately uncorrelated between scales and reduces the impact of boundary conditions (see Section Appendix B). Consecutively, for each of the 8 levels (or wavelet bands) one WV is estimated. Following Abry and Veitch (1998), we perform an analysis of the WV versus scale in a log-log diagram. The same

WV decomposition is applied to the reference time series. Here, we use the one released by Fröhlich (2006), based on its properties in time domain, discussed in the next section. The coefficients of the input time series are then rescaled with the WV ratio between the wavelet coefficients of the two time series. We perform the rescaling on the three first levels of decomposition. Those ones were identified by analysing the previous TSI products and are the ones that are mostly affected by the data fusion process. With our approach, no assumption has to be made about the noise structure, as coming e.g. from a power-law or quantization noise. It can be summarized as adding the right amount of white noise in the right frequency band and does not distort the underlying signal. Finally, we reconstruct the time series by inserting the rescaled coefficient in the inverse of the decomposition function used in the first step called the IMODWT. A comprehensive description of the wavelet filter is given in the appendices.

4 Results and Discussions

4.1 Time-Frequency Analysis of The Composite Time Series

To perform a time-frequency analysis on the 41-year TSI composite, we first produce our time series with the 3-step method. The previous products released by Dudok de Wit et al. (2017), Dewitte and Nevens (2016) and Fröhlich (2006) are called respectively *Composite 1 (C1)*, *Composite 2 (C2)* and *Composite 3 (C3)* in the following text. The new TSI composite is named *Composite PMOD- Data Fusion (CPMDF1)* and after applying the wavelet filter *(CPMDF2)*. Figure 2 display the composite time series. The estimation of the TSI at the solar minimum across the various solar cycles from 1980 to present days are estimated in Table 2.

TSI level ($\mu \pm \sigma$ [W/m^2])		Name of the Composites									
		C1		C2		C3		CPMDF1		CPMDF2	
		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Solar Cycle 21/22	*Minimum* (SM_1)	1360.30	0.14	1362.82	0.12	1360.58	0.12	1360.57	0.11	1360.57	0.12
	$\Delta I_{21/22-20/21}$	-	-	-	-	-	-	-	-	-	-
Solar Cycle 22/23	*Minimum* (SM_2)	1360.68	0.14	1362.90	0.16	1360.57	0.15	1360.56	0.14	1360.56	0.14
	$\Delta I_{22/23-21/22}$	**0.38**	**0.14**	**0.08**	**0.14**	**-0.01**	**0.14**	**-0.01**	**0.11**	**-0.01**	**0.13**
Solar Cycle 23/24	*Minimum* (SM_3)	1360.53	0.04	1362.89	0.04	1360.42	0.06	1360.46	0.05	1360.46	0.05
	$\Delta I_{23/24-22/23}$	**-0.15**	**0.12**	**-0.01**	**0.13**	**-0.15**	**0.12**	**-0.10**	**0.10**	**-0.10**	**0.12**
Solar Cycle 24/25	*Minimum* (SM_4)	-	-	1362.88	0.07	-	-	1360.41	0.08	1360.41	0.09
	$\Delta I_{24/25-23/24}$	-	-	**-0.01**	**0.06**	-	-	**-0.05**	**0.04**	**-0.05**	**0.07**

Table 2: Estimation of TSI at solar minimum (*Minimum*) over last 41 years from the TSI time series (mean μ and standard deviation σ) released by Dudok de Wit et al. (2017) (*C1*), by Dewitte and Nevens (2016) (*C2*) and by Fröhlich (2006) (*C3*). The new TSI composite is abbreviated to (*CPMDF1*) and after using the wavelet filter (*CPMDF2*). The difference of irradiance between solar minima (SM) from consecutive solar cycles (e.g., $\Delta I_{22/23-21/22}$) is also displayed with the uncertainties (bold text)

Note that the solar minima are underlined in Figure 2 (see yellow boxes). The solar minimum periods are chosen according to Dudok de Wit et al. (2017) and Finsterle

et al. (2021) by looking at the lowest value in the yearly-averaged sunspot number and then average the irradiance values over a one-year interval centered on that date. When comparing the mean difference between the product and our time series over the various solar minima, the new composites agree with *C1* at 0.15±0.08 W/m^2, *C2* at 2.37± 0.08 W/m^2, and *C3* at 0.02±0.01 W/m^2. It is the same order of magnitude if we use the wavelet filter (see *CPMDF2*). The difference is marginal for *C1* and *C3*, within the 1-sigma interval of 0.2 W/m^2. This value is defined in Finsterle et al. (2021) based on the inter quantile range estimated on the difference with the reference time series. The large difference with *C2* is due to the absolute level estimated from DIARAD/SOVIM. If we compare with the nominal TSI value of 1361 W/m^2 adopted by the IAU 2015, averaged over Solar cycle 23, the offset is equal to 2.44 W/ m^2. By rescaling *C2* to the nominal TSI value, the discrepancy is reduced to 0.08±0.06 W/m^2, hence within the 1-sigma interval. Figure Appendix C.5 in the appendix shows the box plot of the difference between the products (i.e. *Composite 1, 2, 3*) and the new TSI composite time series. The mean value is less than 0.2 W/ m^2, with the smallest value associated with *C3* around 0.11 W/ m^2. The results are improved at an order of $\sim$ 0.05 W/ m^2 with applying the wavelet filter.

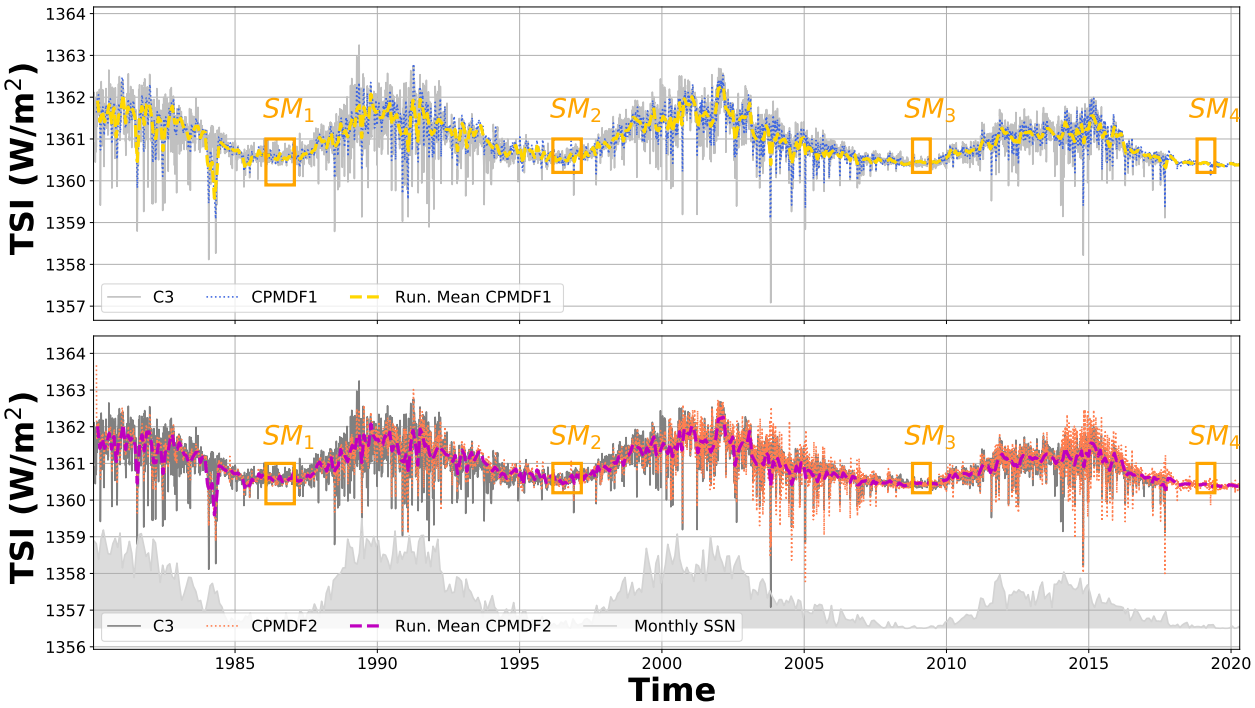

Figure 2: New Compsite (*CPMDF1*, blue) and with wavelet filter *CPMDF2* (orange) based on merging 41 years of TSI measurements. For comparison, *C3* (Fröhlich, 2006) is also shown (grey line). A 30-day running mean of *CPMDF1* is shown as the yellow/purple dash line. The orange boxes are associated with the solar minima (SM) for each solar cycle described in Table 2. For context the monthly sunspot number is also displayed.

Now, we perform a PSD analysis of the composite time series displayed in Figure 3. We can underline the 4 frequencies (11.5 year, 27, 9 and 7 days) related to solar activities and described in the analysis of Fröhlich et al. (1997). The frequency associated with 11.5 years is the Schwabe cycle. The quasi 27-day solar cycle is caused by the sun's differential rotation (presumably first observed by Galileo Galilei or Christoph Scheiner in the first half of the 17th century) (von Savigny et al., 2019).The spectrum is divided in three areas (i.e. box A, B and C) following the previous assumptions on the stochastic properties of the TSI observations (and the solar cycle). The definition of the three areas also follows the description of the photospheric activity. The latter associated with granulations, super-granulations and meso-granulations (Andersen et al., 1994; Fröhlich et al., 1997) generates fluctuations in the TSI at different timescales. Due to our 1 day resolution, frequencies associated with phenomena lasting a few hours or less (i.e. most of the granulations - (Fröhlich et al., 1997)) cannot be observed.

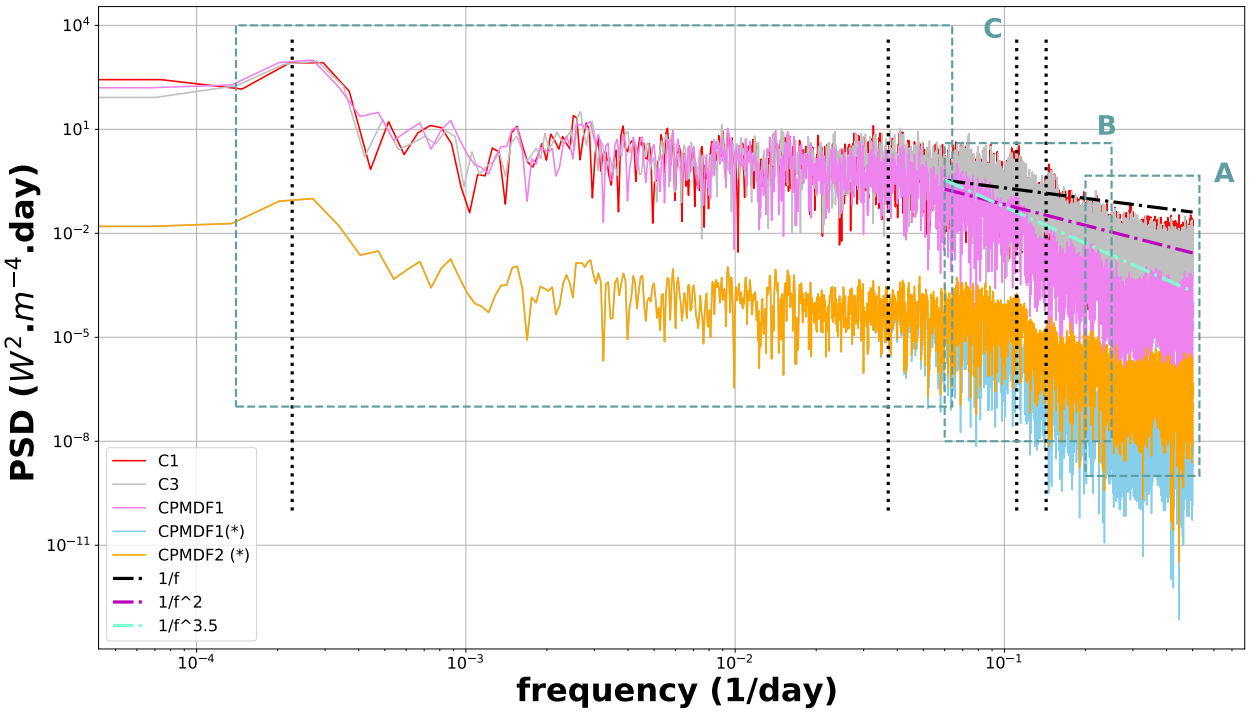

Figure 3: Power Spectrum Density of the TSI *C1* (Dudok de Wit et al., 2017), *C3* (Fröhlich, 2006), together with the new TSI composite produced with the current method *CPMDF1* and applying the wavelet filter *CPMDF2*. The $(*)$ means that the time series are shifted by rescaling the amplitude by -4 $W^2.m^{-4}$.day in the log-log plot. Box A, B and C refer to the different sections of the PSD: A is centered on the high frequency ($\sim$ 3 days) showing the flattening of the PSD; B is the power-law which is mainly due to coloured noise (correlations between 20 and 6 days) within the time series; C emphasises the low frequency associated with the stochastic and deterministic parts of the solar cycle and long-term correlations. The dash lines are the various power-law models when varying the exponent only for indication. The vertical doted lines (black) mark the frequencies at 11.5 years, 27, 9 and 7 days (left to right).

Box *A* shows a flattening of the curve at high frequencies. In the appendices, Figure Appendix C.4 displays the PSD of data recorded by the VIRGO/SOHO radiometer (PMO6V) for the degradation corrected TSI observations on the main channel (VIRGO-A) released in PMO6v21 (Finsterle et al., 2021). The time series is displayed with daily and minute sampling rate. For comparison we also show the PSD of *C3* and *CPMDF2*. The PSD of the products with a daily sampling rate experience the same flattening at high frequencies, whereas the monotonically of the curve disappears in the sub-daily frequency band. Shapiro et al. (2017) discuss that the high frequencies are associated with the radiometer technical characteristics and satellite movements (i.e. open/close shutter, orbit revolutions). Andersen et al. (1994); Fröhlich et al. (1997) show also that the solar noise does flatten in this frequency band. We can then conclude that the flattening is due to the low-sampling rate in the TSI composites.

Box *B* is the power-law or the frequency ramp (between 0.006 and 0.25 day^{-1}). This phenomenon is due to the existence of correlations in the observations. It is arguable that this power-law describes the long-term correlations (i.e. over years), due to the ramp spanning frequencies over only a few days (4 - 20 days). Therefore we can only speculate what underlying process can generate it. For example, it could be an unknown diffusion process associated with the sun's activity which could be modeled with a specific coloured noise called Matérn process. Nonetheless, the steepness of this ramp shows the degree of correlation or the type of stochastic noise within the time series, by fitting a power-law model such as $S(f) \sim 1/f^{\beta}$. The exponent β defines the type of coloured noise: flicker noise corresponds to β equal to 1, a random walk to β equal to 2, and white noise with β equal to 0 (Montillet et al., 2021). Now, Figure 3 displays that the steepness of the ramp is above 2, and even above 3 for the new TSI composite (*CPMDF1*).

Two main factors can explain these results:

i) The high steepness of the ramp in the new TSI composite *CPMDF1*, with β above 3.5 shown in Figure 3, is intrinsic to data fusion. The fusion process behaves as a wide band filtering, similarly to a, e.g., Butterworth filter of low order. If one assumes that the filtered high frequency noise is most likely to be Gaussian, i.e. the slope of the power spectral density in the high frequency domain is close to 0, a smooth filtering will yield a power-law noise. The resulting power-law noise properties depend on the transition band of the filter so that the white noise is transformed into a power-law noise with a variable exponent β. This effect has been exposed in (Accardo et al., 1997) and (Kermarrec, 2020) where high frequency Gaussian noises masked the desired fractional noise. In our composite time series, the power of the bandwidth noise is a nonlinear function of the input parameters used in the fusion process such as the number of inducing points to train the dual kernel. As discussed in Section 3.1, the number of inducing points is an important parameter which drives the quality of the data fusion process. It is recommended to use a large number depending on the size of the series to fuse together, but we are limited to avoid an excessive computing time. Fortunately, the wavelet filter provides an answer to deal with the bandwidth noise. Figure 3 shows the results before and after applying the filter on the frequency spectrum. Also, Figure 2 displays the composite time series with and without applying the wavelet filter. Visually in time, this is similar to having introduced white noise as the time series may appear slightly noisier when using the wavelet filter. That is why the uncertainties associated with the estimated solar minima are slightly higher (~ 0.01 W/ m^2) with the filtered time series (see Table 2). However, from a spectral perspective, this is not the case. The high frequency domain in Box *B* matches well with our reference product (*C3*). We reconstruct the high frequency bandwidth of the fused time series. Figure Appendix C.5 highlights the improvement without modifying the statistical quantity such as median value, upper and lower quantile. Therefore, the wavelet filter improves the 41-year TSI composite time series from data fusion in an effective way without introducing additional computing time.

ii) In the *C1*, *C2* and *C3*, the stochastic noise properties include the correlation from the stochastic part of the solar cycle. In Dudok de Wit et al. (2017), the authors sub-

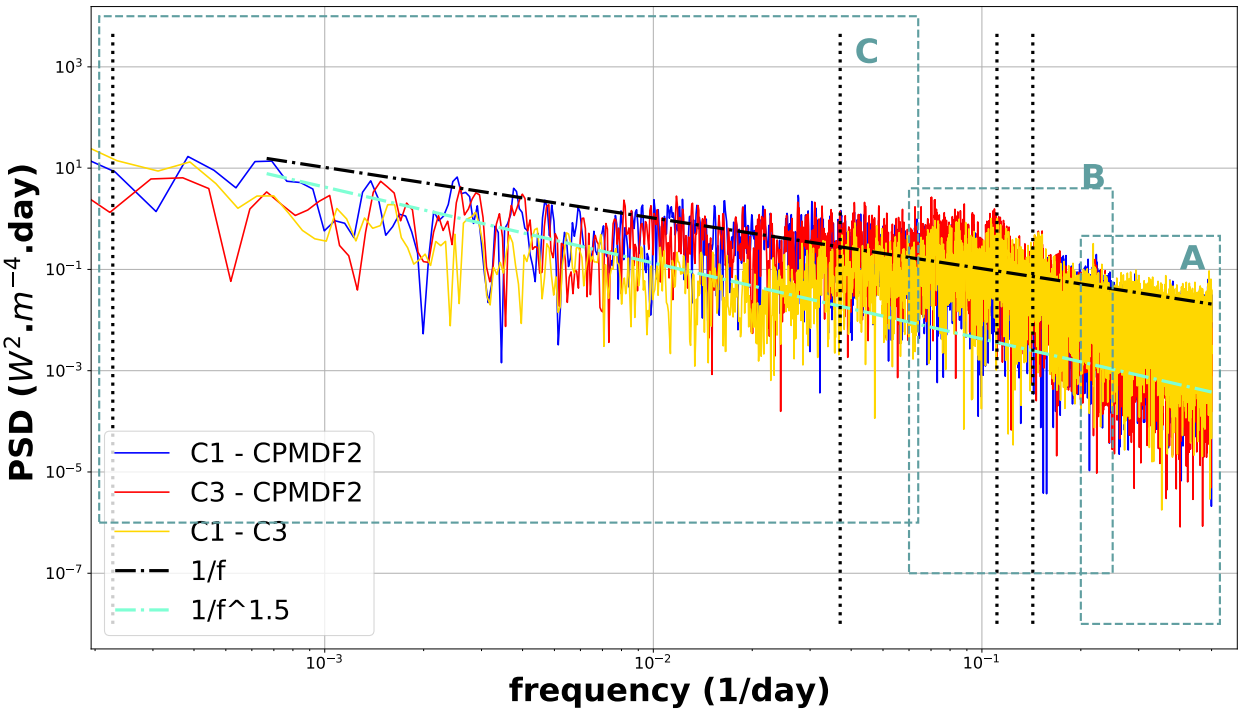

Figure 4: Power Spectrum Density of the difference of composite time series including *C1 - CPMDF2*, *C3 - CPMDF2* , *C1 - C3*. The vertical doted lines (black) mark the frequencies at 11.5 years, 27, 9 and 7 days (left to right).

tracted various TSI time series from different missions (i.e. ACRIM1, ACRIM2, ACRIM3, TIM) in order to eliminate the solar cycle, resulting with only the stochastic properties i.e., a mix between the two instruments. Figure 4 displays the PSD of the difference of 41-year TSI composite. The frequency ramp is mostly attenuated. We can compare its steepness with the various power-law models, hence concluding that the difference composite time series have an exponent β within the interval $]1, 1.5]$. The power-law model is not limited anymore to box B and it includes Box C which advocates for long-term dependencies over years associated with the photospheric activity. This result supports the conclusions in Dudok de Wit et al. (2017).

Finally, box C is associated with the low frequencies (0.006 - 0.00015 day^{-1}). They are assumed to be mostly related to the deterministic part of the solar cycle and the long-term correlations (i.e. lasting up to years). In the appendices, Figure Appendix C.3 shows the spectrum of the *C1* with and without the solar cycle. To remove this cycle, we subtract the time series with a running mean with a 5-day window. We clearly see that the low frequencies in box C have a lowest power in the PSD after subtracting the running mean, hence supporting our assumption. In addition, this frequency band also contains some of the coloured noise linked to long-term correlations (over years). Previously, we have discussed the analysis of Figure 4 when subtracting two TSI composites. We have concluded that the power-law can be extended within Box C highlighting the long-term correlations due to the sun's activity.

4.2 Investigating the solar minimum variations

Once the 41-year TSI composite time series is obtained, we can study the existence of variations in the solar minima. There are two approaches:i) the variations between consecutive solar minima ii) the global fluctuations over the duration of the time series.

The estimation of the variations between two consecutive solar minima is challenging based on the analysis of the PSD of the composite time series. In order to be statistically robust, one needs to take into account the long-term correlations generated by the coloured noise. Therefore, we follow the same methodology as in (Dudok de Wit et al., 2017), where we differentiate the estimated irradiance at solar minima between four consecutive cycles. The results are shown in Table 2 (see e.g., $\Delta I_{22/23-21/22}$). Note that there is an unexpected difference -i.e. about twice the amplitude - for the value of $\Delta I_{22/23-21/22}$ for *C1* between our estimation and the results in Dudok de Wit et al. (2017). Overall, the fluctuations of solar minima between Solar Cycle 21/22 and 22/23 ($\Delta I_{22/23-21/22}$) do not agree between the composites. For example, the difference is positive for the *C1* and *C2*, whereas negative for the other composites. This disagreement can be due to the processing of the TSI observations for the first missions (e.g., HF, ERBE) discussed in Section 3.1. The fluctuations between the other solar cycles (i.e Solar Cycle 22/23, 23/24, 24/25) is more homogeneous between the various composites with an averaged value of -0.10 ± 0.05 and -0.04 ± 0.02 for $\Delta I_{23/24-22/23}$ and $\Delta I_{24/25-23/24}$ respectively. There is a decrease between the differences of solar minima (after cycle 21), which could advocate the presence of a linear trend in the TSI composite. This result is also supported by the fact that the fluctuations between the other cycles (i.e 22/23, 23/24, 24/25) are negative for all the composites. However, this result is downplayed by the large uncertainties associated with the difference between solar minima - up to 10 times the value.

For the study of the global fluctuations, our approach is inspired by the estimation of a tectonic rate in geodetic time series (Davis et al., 2012; Montillet & Bos, 2020). The problem is formulated into a joint estimation of a functional and stochastic models. The functional model is composed of two terms a linear trend and a periodic signal with 4 frequencies (11.5 years, 27, 9 and 7 days) based on our PSD analysis. Because of the flattening experienced by the TSI composite time series at high-frequency (see above discussions - Box A), the use of the General Gauss-Markov model (GGM) with white

noise is appropriate in order to model the frequency ramp feature of the PSD (i.e. Box B). The justification of the model and the estimation of the parameters (using a MLE) are described in the appendices. Now, all the composite time series experience a much lower solar minima at the end of cycle 21 around 1986, it renders the fitting of the deterministic part of the solar cycle difficult with a periodic signal. That is why we perform this study by splitting the composite time series in two time periods 1980-2021 (including cycle 21) and 1987-2021 (starting at cycle 22). The analysis of the functional model fitting the residual shows that the model fits best when using the period 1987- 2021. This confirms our previous study where the exclusion of the solar cycle 21 allows to discute the trend between the difference of solar minima.

TSI level ($\mu \pm \sigma$ [W/(m^{2}yr)])				
Period	1980 - 2021		1987 - 2021	
Amplitude Solar Trend	μ	σ	μ	σ
C1	-0.002	0.005	-0.001	0.007
C2	-0.001	0.006	-0.003	0.007
C3	-0.009	0.006	-0.011	0.009
CPMDF1	-0.015	0.008	-0.010	0.008
CPMDF2	-0.016	0.008	-0.011	0.007

Table 3: Estimation of the linear trend (mean μ and uncertainty σ) via MLE using the GGM model together with white noise for the TSI composite time series released by Dudok de Wit et al. (2017) (*C1*), by Dewitte and Nevens (2016) (*C2*) and by Fröhlich (2006) (*C3*); applying (i.e. *CPMDF2*) or not (i.e. *CPMDF1*) the wavelet filter; and choosing different time periods. yr means year

Table 3 displays the results for each TSI composite time series. *C3* has the largest trend for both periods 1980 - 2021 and 1987 - 2021 compared with *C1* and *C2*. The trend of the new products *CPMDF1* and *CPMDF2* is larger than the previous product for the period 1980 - 2021. However, it is the same order of magnitude as *C3* for the second period.

Averaging the estimated trend for all the previous products produces -0.004 ± 0.004 W/(m$^2 yr$) and -0.005 ± 0.004 W/(m$^2 yr$) for the periods 1980 - 2021 and 1987 - 2021. When we add the new products, the average trend is -0.008 ± 0.006 W/(m$^2 yr$) and -0.012 ± 0.014 W/(m$^2 yr$) for the same periods. The overall estimate, using the results from all the TSI composite time series and both period, is equal to $\sim -0.009 \pm 0.010$ W/(m$^2 yr$). Overall for each product, the uncertainty associated with the estimated trend is large, mostly larger than the amplitude of the trend. This result means that the estimated amplitude is statistically insignificant: the stochastic properties of the composite time series are likely the source of the variations. This result corroborates the previous results based on the estimation of the variations between two consecutive solar minima. Both are showing the same pattern after cycle 21. Note that these conflicting decadal trends exhibited by the previous TSI composites (*C2* and *C3*) are discussed by Yeo et al. (2014) using proxy data. Furthermore, most of the estimated amplitudes are negative. This result has a certain significance related to the solar noise which describes the solar activity, hence meaning that over the last 41 years there has been a slowly decreasing solar activity. This result is supported by several studies focusing on the forecast of the sun's activity over the next 80 years (Steinhilber & Beer, 2013; Velasco Herrera et al., 2015).

5 Conclusions

This work focuses on producing a new 41-year long TSI composite time series using a 3-step method which can be used to study the solar cycle modulation and the Earth's energy budget. We have performed a time-frequency comparison of our new TSI composite with previous releases. The results show that the mean value difference over the solar minima is below the 1 sigma confidence interval of 0.2 W/m^2, i.e. a maximum of 0.15 W/m^2 with *C1* and a minimum of 0.02±0.01 W/m^2 with *C3*. In terms of frequency spectrum comparison, we observe a flattening at high frequencies for all products which is linked to the various instrumental noises and the low sampling rate (1 day). We expect that future TSI composite time series including observations recorded from future missions will be produced with a higher time resolution (i.e. hourly, sub-hourly) in order to include frequency in the meso-granulations, granulations and p-modes frequency bands.

Secondly, the power spectrum experiences a power-law between 4 and 20 days which could correspond to an unknown diffusion process. The steepness of this ramp is more accentuated for our new TSI composite time series than in the previous products. The power-law is between 2.5 and 3 for the previous releases, and between 3 and 3.5 with our new product. This increase steepness is a weakness of the data fusion process, which can smooth the short-term and long-term correlations. It is a nonlinear effect of the input parameters (e.g., inducing points). Increasing this number decreases the steepness to a certain extend. This problem is intractable when using GPs with a very large number of inducing points (e.g., 2000 points) due to computational complexity. Nevertheless, the implementation of a wavelet filter has shown that we can efficiently reconstruct the high frequency bandwidth, hence having a PSD comparable with previous products and without the cost of increasing the processing time. In addition, when removing the solar cycle by differencing two TSI composite time series, a power-law is observed over the whole frequency band. The power-law exponent varies between $[1, 2]$. It highlights the presence of long-term correlations from the solar noise.

Finally, our approach permits the estimation of a trend in the 41-year composite TSI time series which could reflect variations in the solar activity. The analysis of the irradiance difference (ΔI) estimated at two consecutive solar minima in order to detect a trend is inconclusive due to large uncertainties. Our results using a joint inversion of both a functional and stochastic noise models show that the estimated amplitude is below $\sim -0.009 \pm 0.010$ W/(m$^2 yr$) based on the analysis of all the 41-year TSI composite time series used in this study. This number is not statistically robust due to the large uncertainties. Therefore, it is impossible to conclude in the existence of a linear trend in the TSI composite time series. Any visual effects or short-term trends are most likely related to the coloured noise rather than a physical phenomenon generated by the sun's activity corroborating previous discussions (Dudok de Wit & Kopp, 2020) and supporting recent analysis (Schmutz, 2021). Nevertheless, the negative sign from this trend, which is also shown in the analysis of ΔI, indicates a slowly decreasing solar activity in the last 41 years.

Acknowledgments
We acknowledge the life-long dedication to the TSI community of the late Dr. Claus Fröhlich in memoriam as former director of PMOD/WRC, PI of SOHO/VIRGO and his invaluable contribution to the analysis of TSI observations. Among them, he produced the first reconstruction of the TSI composite in 2006 which is also used in this work (i.e. *C3*). The TSI composite is retrieved from the PMOD archive and it is available on demand to Dr. W. Finsterle. The TSI composite *C1* is available for download at *http* : *//www.issibern.ch/teams/solarirradiance*. *C2* is available on request to Dr. S. Dewitte. The new composites *CPMDF1* and *CPMDF2* should be accecible on *ftp* : *//ftp.pmodwrc.ch/pub/data /irradiance/*. The data related to the monthly/daily mean sunspot numbers are retrieved from *http* :

//www.sidc.be/ silso/datafiles. *TIM /SORCE* and *TIM/TSIS* time series are downloaded from *https* : *//lasp.colorado.edu/home/sorce/data/tsi−data* and *ACRIM* from *www.acrim.com*. *PMODv21a* is available at *ftp* : *//ftp.pmodwrc.ch/pub/data /irradiance/virgo/TSI/*. *PREMOS (v1)* can be accessed at *http* : *//idoc−picard.ias.u−psud.fr/sitools/client−user/Picard/project−index.html*. Dr. J.-P. Montillet, Dr. W. Finsterle, Dr.Haberreiter and Prof. Schmutz gratefully acknowledge the support of Karbacher-Funds. Dr.G. Kermarrec would like to acknowledge the Deutsche Forschungsgemeinschaft under the project KE2453/2-1 which permitted the development of the wavelet filter to analyze correlated noise, opening the door for further studies related to laser observations.

Appendix A Model Descriptions for the Estimation of the Linear Trend

Following the discussion in Section 4, the stochastic noise model of the TSI time series is described with the variance (Williams et al., 2004):

$$E\{\psi^T\psi\} = \sigma_{wn}^2\mathbf{I} + \sigma_{pl}^2\mathbf{J}(\beta) \tag{A1}$$

where the vector $\psi = [\psi(t_1), \psi(t_2), ..., \psi(t_L)]$ is a multivariate continuous-time stochastic process. At each time step, we define $\psi(t_i) = \psi_{wn}(t_i) + \psi_{pl}(t_i)$, with $\psi_{wn}(t_i)$ and $\psi_{pl}(t_i)$ the white Gaussian noise (zero mean) and the coloured noise (or power-law noise) sample respectively. T is the transposition operator, $\mathbf{I}$ the identity matrix, σ_{pl}^2 the variance of the power-law noise and $\mathbf{J}(\beta)$ the covariance matrix of the power-law noise ($\beta > 0$). The definition of $\mathbf{J}$ depends on the assumptions on the type of coloured noise (e.g., Flicker, random-walk).

The functional model $s_0(t)$ (at epoch t) is based on the polynomial trigonometric method (Williams et al., 2004; Montillet & Bos, 2020).

$$s_0(t) = at + b + \sum_{j=1}^{N}(G_j\cos(D_j t) + E_j\sin(D_j t)) \tag{A2}$$

with a and b the coefficients of the linear rate; the deterministic part of the solar cycle is modeled by a sum of cos and sin functions with coefficients G_j and E_j. Note that D_j is equal to $2\pi fq_j$, and fq_j are different frequencies (e.g., 11.5 years, 27, 9 and 7 days) which are determined by analysing the frequency spectrum of the TSI composite time series (see Section 4). We perform a joint estimation of the functional and stochastic models based on a maximum likelihood estimator (MLE). To recall (Bos et al., 2020), the log-likelihood for a time series of length n can be rewritten as:

$$\ln(Lo) = -\frac{1}{2}\left[n\ln(2\pi) + \ln(\det(\vec{C})) + (\vec{x}_0 - \vec{A}\vec{z})^T\vec{C}^{-1}(\vec{x}_0 - \vec{A}\vec{z})\right] \tag{A3}$$

This function must be maximised. Assuming that the covariance matrix $\vec{C}$ is known, then it is a constant and does not influence finding the maximum. $\vec{C}$ is here defined by Eq. (A1). The term $(\vec{x}_0 - \vec{A}\vec{z})$ represents the TSI observations minus the fitted model. Note that $(\vec{A}\vec{z})$ is the matrix notation of s_0. The last term can be written as $\vec{x}^T\vec{C}^{-1}\vec{x}$ and it is a quadratic function, weighted by the inverse of matrix $\vec{C}$. To select the functional model of the solar signal, and therefore estimate the associated parameters, we have formulated the assumptions in Section 3 and the time-frequency analysis in Section 4. The value of n is here equal to the number of observations in the TSI composite time series ($\sim$ 15330 observations).

For an optimal estimation of the parameters with the MLE, a strong assumption in geodetic time series is the so-called Gauss-Markov hypothesis (e.g., (Bos et al., 2020)) which states that the noise is Gaussian distributed. With the TSI time series, the functional model cannot remove completely the stochastic part of the solar cycle. Therefore, the distribution of the residual time series cannot be guaranteed to be a perfect Gaussian. We then formulate two important assumptions in order to get an unbiased estimate. First, the mean of the coloured noise is slowly varying with time, which means that a/ the noise is non-stationary around the mean, and b/ no intermittency in the time series, i.e. no events creating short high bursts or sudden large deviations for the mean (e.g., aggregations). Secondly, the noise is assumed to be homoscedastic which means there is no change of variance across the observations in the 41-year composite time series (Montillet et al., 2021).

The particularity of the TSI composite frequency spectrum has been discussed above, and in particular its flattening over the high frequencies. Therefore, a simple power-law

noise model as described in Section 3 with the covariance $\mathbf{J}(\beta)$ is not appropriated for our ML estimation. Instead, we use the Generalized Gauss-Markov (GGM) noise model which has the advantage to flatten in high frequency. The power spectrum of the GGM noise is defined by Bos et al. (2020) such as:

$$S(f) = \frac{2\sigma^2}{f_s^2}\left[1 + \phi^2 - 2\phi\cos\left(2\pi\frac{f}{f_s}\right)\right]^{-\beta/2} \tag{A4}$$

where ϕ is an important parameter to decide when the flattening occurs in the PSD. In our study with TSI time series, we have fixed ϕ to 1.0699. Also from Eq. (A4), if ϕ equal to 1 the PSD is similar to an approximation of the power-law model. For more information and discussions about this model, we invite the reader to refer to Bos et al. (2014). Note that we use the Hector package to do the joint model estimation.

Appendix B Description of the Wavelet Filter

This section describes comprehensively the wavelet filter discussed in Section 3.3. Figure Appendix B.1 underlines the description of the wavelet filter below.

B1 Methodology

Our algorithm is based on an unbiased wavelet-based estimator developed by McCoy and Walden (1996) and Abry and Veitch (1998) for the correlation analysis and correction of some of the long-range dependency within the TSI observations. Our methodology relies on the decomposition of the variance of a time series on a scale by scale basis, the so-called WV, which can be simply interpreted as the variance of a process after filtering by a wavelet bandpass filters (Percival & Guttorp, 1994).

Fractal processes have a power of the form $1/f^{\beta}$ for a range of frequency f close to 0. The WV multiscale analysis provides an efficient and highly robust estimator of the fractal parameter of a process against, e.g., the addition of a deterministic trend or sinusoidal pattern. The WV versus its scales have a logarithmic linear relationship. The slope in a log_2-diagram is related to the power law of the process and can be estimated by ordinary least-squares (Abry & Veitch, 1998). Here, we propose to use the WV to eliminate the correlations induced at high frequency by the windowing used to fuse TSI from different dataset. The procedure is inspired from the work of Guerrier et al. (2013) on composite stochastic processes and is based on the standardized distance between the WV of a reference process and the one to analyse.

B2 The MODWT

By design, the wavelet's advantage is their ability to simultaneously localize a process in time and scale. Low scales are related to long periodic behaviour, whereas high scales focus on brief phenomena, i.e. high frequencies. The wavelet transformation breaks the time series down into a scaled and shifted version of the mother wavelet. Unfortunately, the sample size of the orthogonal discrete wavelet transform (DWT) is limited to a power of 2 so that the number of scaling and wavelet coefficients at each level of resolution decreases by the same factor. This results in a loss of information as well as the introduction of ambiguities in the time domain. MODWT (Cornish et al., 2006) carries out the same steps as the DWT without a sub-sampling process. The coefficients of decomposed components in each layer have the same length as the original time-series. This property is favourable for information extraction such as in our case. Mathematically, the MODWT is a convolution operation that can be formulated as circular filter operations of the original time series using the 2 quadrature mirror filters, where $\{\tilde{g}_{J,l}\}$ (with $l = [0, .., L_j]$) represents the single level smoothing filter length L_j and $\{\tilde{h}_{j,l}\}$ (with $l = [0, .., L_j]$) the level-j wavelet filters, i.e. the coefficients of a bandpass filter. The MODWT algorithm is described in Figure Appendix B.1.A. If $c_{0,f} = x_f$ is the input time series of length n, we generate the wavelet coefficients $\{d_{j,f}\}$ and the scaling coefficients $\{c_{j,f}\}$ from $\{c_{j,f-1}\}$, where for a level $j = [0, .., J]$ we have $d_{j,f} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} x_{f-l(mod.n)}$, and $c_{J,n} = \sum_{l=0}^{L_j-1} \tilde{g}_{J,l} x_{n-l(modN)}$ $(n = [0, .., N-1])$. The coefficients $\{d_{j,f}\}$ are the difference between generalized averages of the time series each occupying a width of 2^{j-1}. Therefore, the decomposition at each scale can be understood as a bandwidth filtering of the original time series in the frequency domain with various high and low-pass filters (Cornish et al., 2006). MODWT is known as a shift-invariant wavelet transform. It is a highly redundant version of DWT and is considered ideal for time-series analysis, as it accommodate any sample size. Note that we use the MATLAB wavelet package from the Mathworks ($https://ch.mathworks.com/$) where the MODWT is included.

B3 Choice of the Type of Wavelet and Width, and Wavelet Variance

Using the MODWT for WV requires the choice of the wavelet filter. Here, we follow the work of Cornish et al. (2006) who proposed to use the least asymmetric (LA) wavelets. They exhibit near symmetry about the filter midpoint. This property allows a good alignment of the reconstruction with the original time series by circularly sifting the coefficients. Empirical investigations of the TSI observations have shown that LA(4) was optimal for reconstruction. We then chose to decompose the signal in a 8 levels which we justify by our specific focus on the high frequency domain. The low frequencies do not need particular attention and should be kept intact with the purpose of not losing information from the data fusion.

The WV is built using the wavelet coefficients issued from the MODWT. An unbiased estimator of the WV at scale j is given by $\nu_j^2 = \frac{1}{L_j} \sum_{l=0}^{L_j-1} d_{j,l}^2$, which is a consistent estimator of the true variance as long as a large number of wavelet coefficients are available. For a fractal process, Abry and Veitch (1998) demonstrated that the WV follows a linear relationship in logarithmic scale between ν_j^2 and β. It then allows the estimation of β with a simple ordinary least-squares.

When dealing with composite stochastic process - defined as a mix of bandwidth (fractal) noises - the scales at which a given noise is present have to be identified in a first step (e.g., -4 slope, flicker noise, random-walk noise). An alignment of the WV versus a given number of scales is linked with a bandwidth (fractal) noise. Here, we want to identify the correlated noise introduced by data fusion, which was shown to be found in the high frequency domain. We illustrate our decorrelation procedure using the *C3*. Figure Appendix B.1.C is a log_2-diagram showing the WV versus scales. Our reference time series is marked with the blue dots, together with the fused time series called *CPMDF2* with pink dots. The WV of both time series lies on a straight line. For the reference time series, the slope corresponds approximately to a coefficient equal to -3 in the WV spectrum. The first WV contains most probably an additional WN component as it is slightly over the line drawn from scale 2 and 3. From scale 4, the WV spectrum changes clearly its shape, which corresponds to the Matérn process (saturation at low frequency). Our analysis shows that the high frequency correlated noise is present between the scales 1 and 3.

Having identified the high frequency correlations in the 3 first scales, we propose to simply rescale the WV of *CPMDF1* in order to fit the reference one. The resulting WV spectrum is shown in the green dots. Finally, the filtered time series is obtained using the inverse function inserting our new WV values. Note that the function is the IMODWT, also included in the MATLAB wavelet package. Our new time series is comparable to both *C1*, *C2* (after shifting of ~ -2.44 W/m^2), *C3* and *CPMDF1*, well within the interval of confidence of the reconstruction (0.3 W/m^2) as discussed in Section 4. Note that the MODWT and IMODWT require $O(n.log_2.n)$ multiplications (Percival & Walden, 2000).

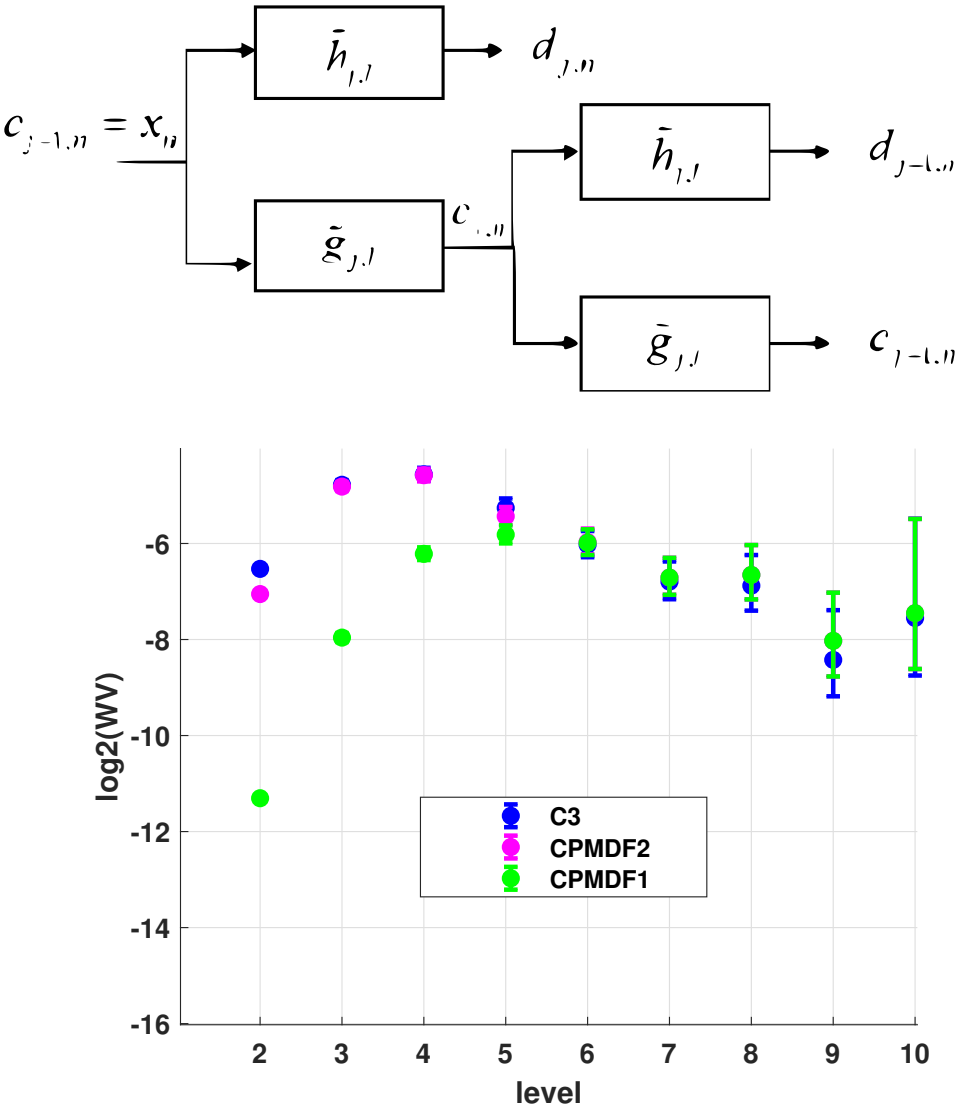

Figure Appendix B.1: (Top) Wavelet filter with the estimation of the coefficients for the MODWT algorithm. (Bottom) Time in Julian Day (JD) WV decomposition of *C3* (blue), *CPMDF1* (green) and the filtered time series (pink).

Appendix C Additional Figures

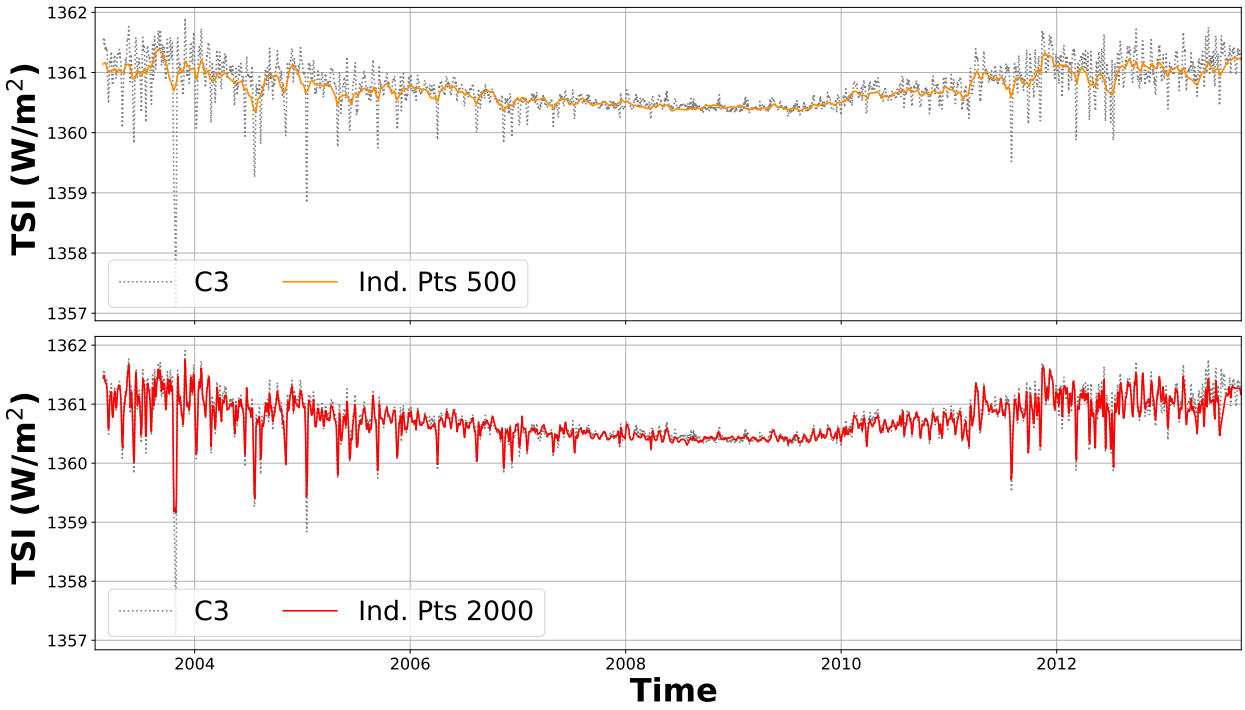

Figure Appendix C.1: Time series of *C3* (Fröhlich, 2006) and the sub-time series in Box 8 (see Figure 1) fusing VIRGO/SOHO, ACRIM 3 and TIM/SORCE using various numbers of inducing points (500, 2000). The sub-time series are aligned on *C3*.

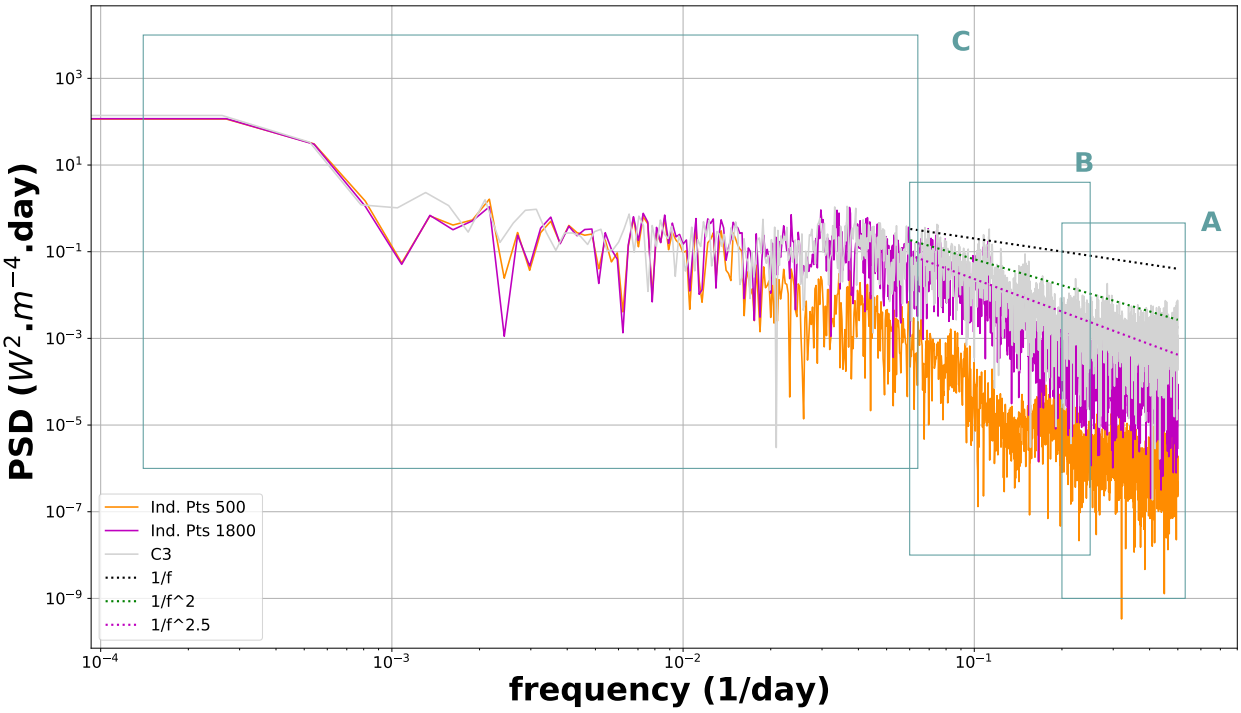

Figure Appendix C.2: Power Spectrum of the TSI *C3* (Fröhlich, 2006) and the sub-time series in Box 8 fusing VIRGO/SOHO, ACRIM 3 and TIM/SORCE using various numbers of inducing points (500, 1800). Box A, B and C refer to the different sections of the PSD: A is centered on the high frequency ($\sim$ 3 days) showing the flattening of the PSD; B is the power-law which is mainly due to coloured noise (correlations between 20 and 6 days) within the time series; C emphasises the low frequency associated with the stochastic and deterministic parts of the solar cycle and long-term correlations.The dash line is the power-law model when varying the exponent only for indication.

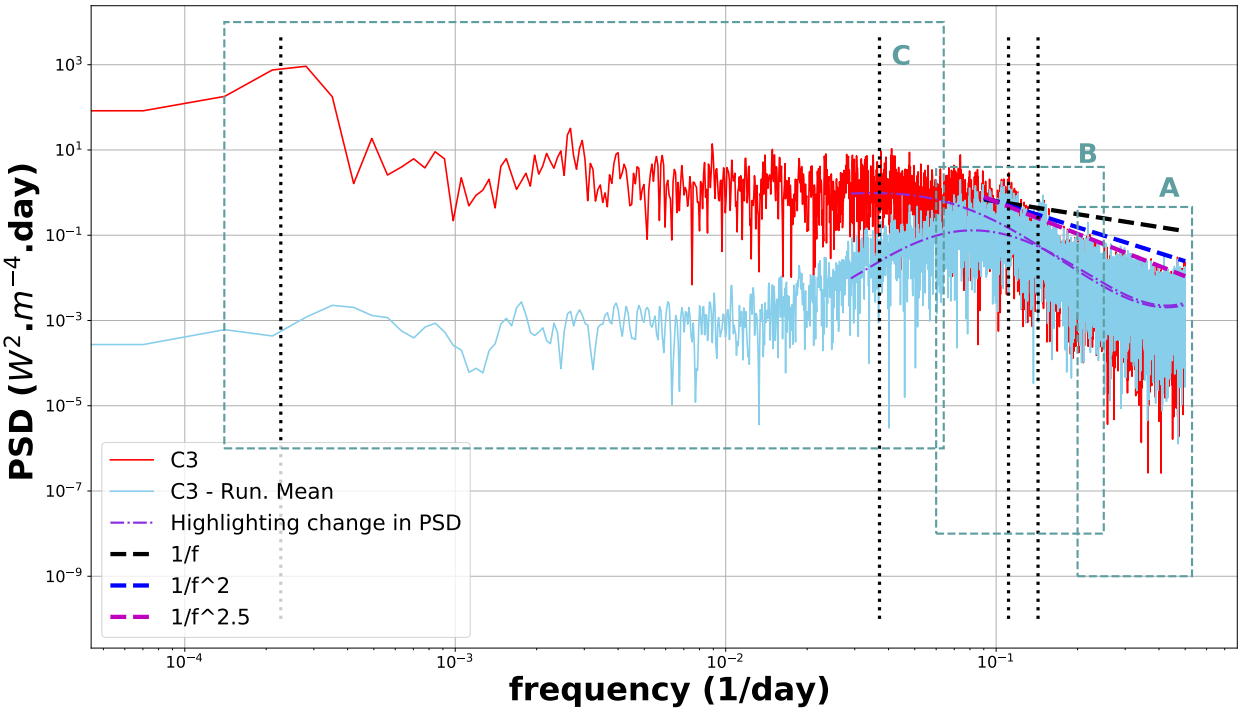

Figure Appendix C.3: Power Spectrum of the TSI *C3* (Fröhlich, 2006) with and without removing partially the solar cycle via a running mean (with a 5 day window). Box *A*, *B* and *C* refer to the different sections of the PSD. The dash-doted lines are the various power-law models when varying the exponent. The vertical dotted lines (black) mark the frequencies at 11.5 years, 27, 9 and 7 days. The purple dash-dot lines are highlighting the change of power in the box *C* before and after removing the solar cycle.

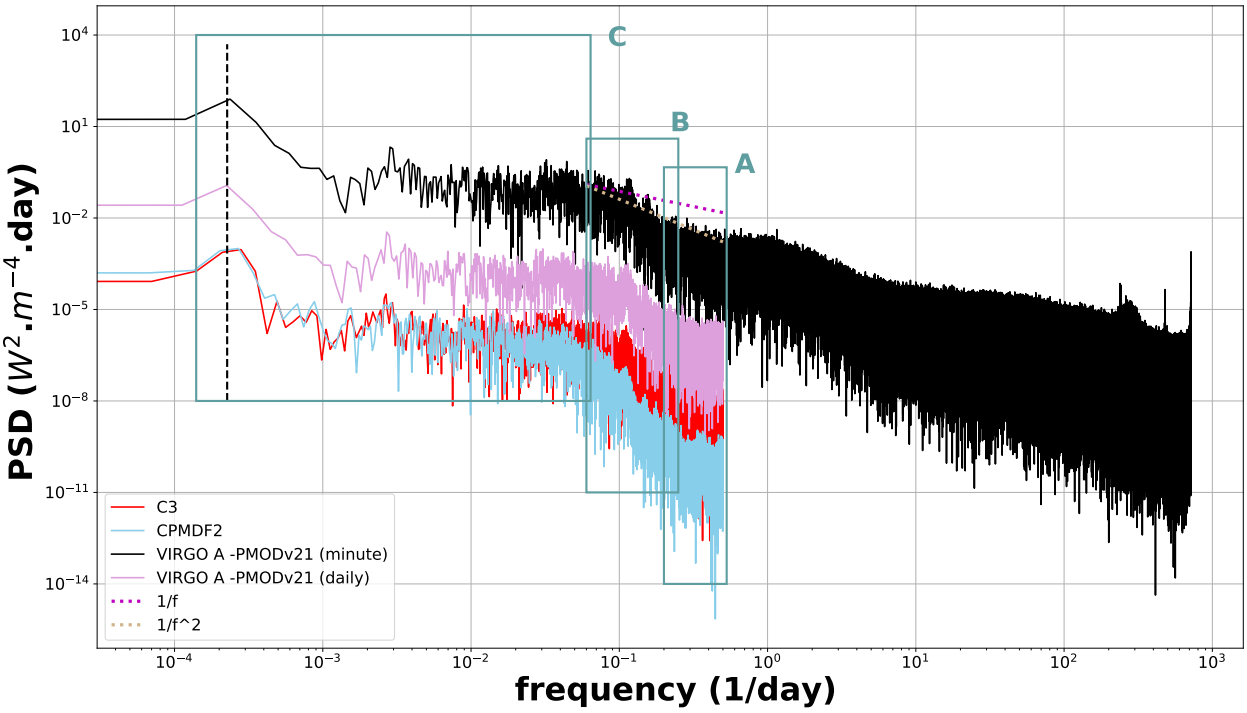

Figure Appendix C.4: Power Spectrum of various products: VIRGO/SOHO with degradation correction (VIRGO-A) from PMO6v21 with minute and daily sampling rate. *C3* is the TSI composite produced by Fröhlich (2006). *CPMDF2* is our new TSI composite including the wavelet filter. Note that for clarity the different spectra have been rescaled by multipying them with -6, -6, -4 and -8.5 $W^2.m^{-4}$.day following the order of the legends from the top. Box *A*, *B* and *C* refer to the different sections of the PSD. The dash lines are the various power-law models when varying the exponent only for indication. The vertical dash line emphasizes the 11.5 year peak.

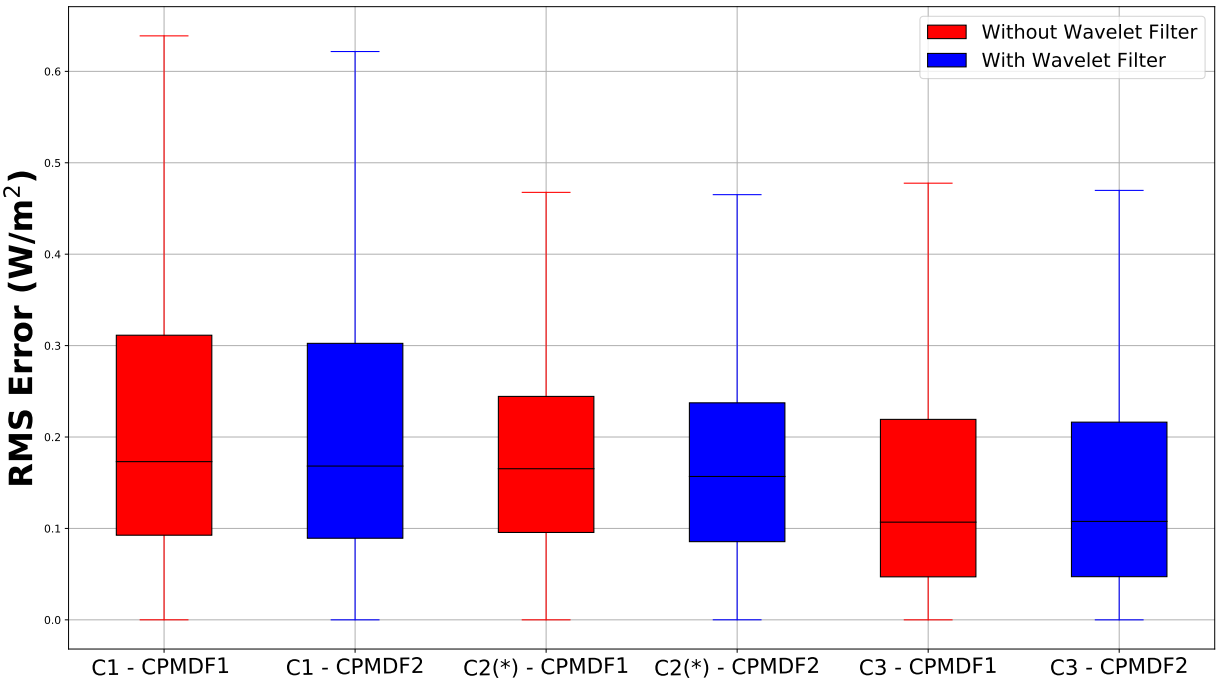

Figure Appendix C.5: Box plot figure comparing the difference between the TSI *C1* (Dudok de Wit et al., 2017), *C2* (Dewitte & Nevens, 2016), *C3* (Fröhlich, 2006) and the new TSI composite produced with the current method in terms of root mean square (RMS) error without (*CPMDF1*) and with the wavelet filter (*CPMDF2*). (∗) means that *C2* is shifted of -2.44 W/ m^2.

References

Abry, P., & Veitch, D. (1998). Wavelet Analysis of Long-Range Dependent Traffic. *IEEE Trans. on Info. Theory*, *44*, 2–15. doi: 10.1109/18.650984

Accardo, A., Affinito, M., Carrozzi, M., & Bouquet, F. (1997). Use of the fractal dimension for the analysis of electroencephalographic time series. *Biol. Cybern.*, *77*, 339–350. doi: 10.1007/s004220050394

Andersen, B. N., Leifsen, T., & Toutain, T. (1994). Solar noise simulations in irradiance. *Sol. Phys.*, *152*, 247–252. doi: 10.1007/BF01473211

Bauer, M., van der Wilk, M., & E., R. C. (2016). Understanding Probabilistic Sparse Gaussian Process Approximations. In *Nips* (Vol. 2016). doi: arXiv:1606.04820

Bos, M., Williams, S. D. P., Araújo, I., & Bastos, L. (2014). The effect of temporal correlated noise on the sea level rate and acceleration uncertainty. *Geophys. J. Int.*, *96*, 1423–1430. doi: 10.1093/gji/ggt481

Bos, M. S., Montillet, J.-P., Williams, S. D. P., & Fernandes, R. M. S. (2020). Introduction to geodetic time series analysis. In J.-P. Montillet & M. S. Bos (Eds.), *Geodetic time series analysis in earth sciences* (pp. 29–52). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-21718-1_2` doi: 10.1007/978-3-030-21718-1_2

Cocchi, M. (2019). *Data fusion methodology and applications* (Vol. 31). Elsevier. doi: 10.1016/B978-0-444-63984-4.21001-6

Cornish, C. R., Bretherton, C. S., & Percival, D. B. (2006). Maximal Overlap Wavelet Statistical Analysis With Application to Atmospheric Turbulence. *Boundary-Layer Meteorol.*, *119*, 339–374. doi: 10.1007/s10546-005-9011-y

Davis, J., Wernicke, B., & Tamisiea, M. (2012). On seasonal signals in geodetic time series. *J. Geophys. Res*, *117*. doi: 10.1029/2011JB008690

Dewitte, S., & Nevens, S. (2016). The Total Solar Irradiance Climate Data Record. *The Astrophysical Journal*, *830*(1), 25. doi: 10.3847/0004-637x/830/1/25

Dudok de Wit, T., & Kopp, G. (2020). *1/f noise in irradiance records affects our understanding of trends in solar radiative forcing.* (AGU conference, A237-08)

Dudok de Wit, T., Kopp, G., Fröhlich, C., & Schöll, M. (2017). Methodology to create a new total solar irradiance record: Making a composite out of multiple data records. *Geophys. Res. Lett.*, *44*, 1196–1203. doi: 10.1002/2016GL071866

Egorova, T., Rozanov, E., Arsenovic, P., Peter, T., & Schmutz, W. (2018). Contributions of natural and anthropogenic forcing agents to the early 20th century warming. *Frontiers in Earth Science*, *6*, 206. Retrieved from `https://www.frontiersin.org/article/10.3389/feart.2018.00206` doi: 10.3389/feart.2018.00206

Feynman, J. (1982). Geomagnetic and solar wind cycles, 1900–1975. *Journal of Geophysical Research: Space Physics*, *87*(A8), 6153-6162. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JA087iA08p06153` doi: https://doi.org/10.1029/JA087iA08p06153

Finsterle, W., Montillet, J., Schmutz, W., Sikonja, R., Kolar, L., & Treven, L. (2021). The total solar irradiance during the recent solar minimum period measured by SOHO/VIRGO. *Scientific Reports*, *11*(7835), 10. doi: 10.1038/s41598-021-87108-y

Fontenla, J., White, O., Fox, P., Avrett, E., & Kurucz, R. L. (1999). Calculation of solar irradiances. i. synthesis of the solar spectrum. *The Astrophysical Journal*, *518*. doi: 10.1086/307258

Fröhlich, C. (2006). Solar Irradiance Variability Since 1978. Revision of the PMOD Composite during Solar Cycle 21. *Space Science Reviews*, *125*, 53–65. doi: 10.1007/s11214-006-9046-5

Fröhlich, C., Andersen, B., & Appourchaux, T. e. a. (1997). First Results from VIRGO, the Experiment for Helioseismology and Solar Irradiance monitoring

on SOHO. *Solar Physics*, *170*, 1–25. doi: 10.1023/A:1004969622753

Fröhlich, C. (2006). Solar irradiance variability since 1978. *Space Sci. Rev.*, *125*, 53–65. doi: 10.1007/s1121400690465

Fröhlich, C., & Lean, J. (2004). Solar radiative output and its variability: Evidence and mechanisms. *Astron. Astrophys. Rev.*, *12*, 273—320. doi: 10.1007/s00159-004-0024-1

Guerrier, S., Skaloud, J., Stebler, Y., & Victoria-Feser, M.-P. (2013). Wavelet-Variance-Based Estimation for Composite Stochastic Processes. *Journal of the American Statistical Association*, *108*. doi: 10.1080/01621459.2013.799920

Haykin, S. (2004). *Adaptive Filter Theory.* Prentice Hall Information and System Sciences Series, Prentic-Hall, Inc.

Kermarrec, G. (2020). On estimating the hurst parameter from least-squares residuals. case study: Correlated terrestrial laser scanner range noise. *Mathematics*, *8*(5). Retrieved from `https://www.mdpi.com/2227-7390/8/5/674` doi: 10.3390/math8050674

Kolar, L., Šikonja, R., & Treven, L. (2020). Iterative Correction of Sensor Degradation and a Bayesian Multi-Sensor Data Fusion Method. *ArXiv*. Retrieved from `https://arxiv.org/pdf/2009.03091.pdf` doi: arXiv:2009.03091

Kopp, G. (2016). Solar Variability Magnitudes and Timescales. *Space Science Reviews*, *6*. doi: 10.1051/swsc/2016025

Kopp, G., & Lean, J. L. (2011). A new, lower value of total solar irradiance: Evidence and climate significance. *Geophys. Res. Lett.*, *38*. doi: 10.1029/2010GL045777

Kren, A. (2015). *Investigating the role of the Sun, the quasi-biennial oscillation, and the pacific decadal oscillation on decadal climate variability of the stratosphere* (Unpublished doctoral dissertation). University of Colorado at Boulder.

McCoy, E., & Walden, A. (1996). Wavelet Analysis and Synthesis of Stationary Long-Memory Processes. *Journal of Computational and Graphical Statistics*, *5*(1), 26–56. doi: 10.2307/1390751

Mekaoui, S., & Dewitte, S. (2008). Total solar irradiance measurement and modelling during cycle 23. *Sol. Phys.*, *247*, 203–216. doi: 10.1007/s11207-007-9070-y

Montillet, J.-P., & Bos, M. (2020). *Geodetic Time Series Analysis in Earth Sciences*. Springer International Publishing.

Montillet, J.-P., He, X., Yu, K., & Xiong, C. (2021). Application of Lévy processes in modelling (geodetic) time series with mixed spectra. *Nonlin. Processes Geophys.*, *28*, 121–134. doi: 10.5194/npg-28-121-2021

Percival, D., & Walden, A. (2000). *Wavelet methods for time series analysis.* Cambridge University Press, Cambridge UK. doi: 10.1017/CBO9780511841040

Percival, D. B., & Guttorp, P. (1994). Long-memory processes, the allan variance and wavelets. In E. Foufoula-Georgiou & P. Kumar (Eds.), *Wavelets in geophysics.* Elsevier.

Prša, A., Harmanec, P., Torres, G., Mamajek, E., Asplund, M., Capitaine, N., ... Stewart, S. G. (2016). Nominal Values for Selected Solar and Planetary Quantities: IAU 2015 Resolution B3. *Astronomical Journal*, *152*(2), 41. doi: 10.3847/0004-6256/152/2/41

Rind, D. H., Lean, J. L., & Jonas, J. (2014). The impact of different absolute solar irradiance values on current climate model simulations. *Journal of Climate*, *27*(3), 1100 - 1120. Retrieved from `https://journals.ametsoc.org/view/journals/clim/27/3/jcli-d-13-00136.1.xml` doi: 10.1175/JCLI-D-13-00136.1

Scafetta, N., Wilson, R., Lee, J., & D.L., W. (2020). *Modeling Quiet Solar Luminosity Variability from TSI Satellite Measurements and Proxy Models from 1980 to 2018.* (AGU conference, A237-07)

Schmutz, W., Fehlmann, A., Finsterle, W., Kopp, G., & Thuillier, G. (2013). Total

solar irradiance measurements with PREMOS/PICARD. In *American institute of physics conference series* (Vol. 1531, pp. 624–627). doi: 10.1063/1.4804847

Schmutz, W. K. (2021). Changes in the total solar irradiance and climatic effects. *J. Space Weather Space Clim.*, *11*, 40. Retrieved from https://doi.org/10.1051/swsc/2021016 doi: 10.1051/swsc/2021016

Shapiro, A., Solanki, S., Krivova, N., Cameron, R., Yeo, K., & Schmutz, W. (2017). The nature of solar brightness variations. *Nat. Astron.*, *1*, 612–616. doi: 10.1038/s41550-017-0217-y

Steinhilber, F., & Beer, J. (2013). Prediction of solar activity for the next 500 years. *Journal of Geophysical Research: Space Physics*, *118*(5), 1861-1867. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgra.50210 doi: https://doi.org/10.1002/jgra.50210

Velasco Herrera, V., Mendoza, B., & Velasco Herrera, G. (2015). Reconstruction and prediction of the total solar irradiance: From the medieval warm period to the 21st century. *New Astronomy*, *34*, 221-233. Retrieved from https://www.sciencedirect.com/science/article/pii/S1384107614001080 doi: https://doi.org/10.1016/j.newast.2014.07.009

von Savigny, C., Peters, D. H. W., & Entzian, G. (2019). Solar 27-day signatures in standard phase height measurements above central Europe. *Atmos. Chem. Phys.*, *19*. doi: 10.5194/acp-19-2079-2019

Williams, S. D. P., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R. M., Prawirodirdjo, L., ... Johnson, D. J. (2004). Error analysis of continuous gps position time series. *Journal of Geophysical Research: Solid Earth*, *109*(B3). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003JB002741 doi: https://doi.org/10.1029/2003JB002741

Willson, R. (1997). Total solar irradiance trend during solar cycles 21 and 22. *Science*, *277*, 1963–1965. doi: 10.1126/science.277.5334.1963

Xiang, N. B. (2019). Revisiting the Question: The Cause of the Solar Cycle Variation of Total Solar Irradiance. *Advances in Astronomy*, *2019*. doi: 10.1155/2019/3641204

Yeo, K. L., Krivova, N. A., & Solanki, S. K. (2014). Solar cycle variation in solar irradiance. *Space Science Reviews*, *136*, 137-167. Retrieved from https://doi.org/10.1007/s11214-014-0061-7 doi: 10.1007/s11214-014-0061-7

Yeo, K. L., Solanki, S. K., Krivova, N. A., & Jiang, J. (2021). The relationship between bipolar magnetic regions and their sunspots. *A&A*, *654*, A28. Retrieved from https://doi.org/10.1051/0004-6361/202141336 doi: 10.1051/0004-6361/202141336

Yeo, K. L., Solanki, S. K., M., N. C., Beeck, B., Unruh, Y. C., & Krivova, N. A. (2017). Solar irradiance variability is caused by the magnetic activity on the solar surface. *Physical review letters*, *119 9*, 091102.