

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

J. -P. Montillet, W. Finsterle, G. Kermarrec, R. Sikonja, M. Haberreiter, W. Schmutz, Thierry Dudok de Wit

▶ **To cite this version:**

J. -P. Montillet, W. Finsterle, G. Kermarrec, R. Sikonja, M. Haberreiter, et al.. Data Fusion of Total Solar Irradiance Composite Time Series Using 41 Years of Satellite Measurements. Journal of Geophysical Research: Atmospheres, 2022, 127, 10.1029/2021JD036146 . insu-03857164

HAL Id: insu-03857164
https://insu.hal.science/insu-03857164

Submitted on 31 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Check for updates

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2021JD036146

Special Section:
Monitoring the Earth radiation budget and its implication to climate simulations: Recent Advances and Discussions

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], and T. Dudok de Wit[4]

[1]Physikalisch-Meteorologisches Observatorium Davos/World Radiation Center (PMOD/WRC), Davos, Switzerland, [2]Institute of Meteorology and Climatology, Leibniz University 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

Abstract Since the late 1970s, successive satellite missions have been monitoring the sun's activity and recording the total solar irradiance (TSI). Some of these measurements have lasted for more than a decade. In order to obtain a seamless record whose duration exceeds that of the individual instruments, the time series have to be merged. Climate models can be better validated using such long TSI time series which can also help to 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 scaled at the nominal TSI value of ∼1361 W/m^2, the difference is below 0.2 W/m^2 in terms of solar minima. Next, we model the frequency spectrum of this 41-year TSI composite time series with a Generalized Gauss-Markov model to help describe 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 $\sim -0.004 \pm 0.004$ 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 artifact due to the colored noise.

Key Points:

- Application of data fusion to merge 41 years of satellite observations into a composite total solar irradiance time series
- A comprehensive time-frequency analysis to characterize the solar cycle and the stochastic noise
- Investigation of variations at solar minima to distinguish between stochastic noise and underlying phenomena linked to the solar activity

Correspondence to:
J.-P. Montillet,
jean-philippe.montillet@pmodwrc.ch

Citation:
Montillet, J.-P., Finsterle, W., Kermarrec, G., Sikonja, R., Haberreiter, M., Schmutz, W., & Dudok de Wit, T. (2022). Data fusion of total solar irradiance composite time series using 41 years of satellite measurements. *Journal of Geophysical Research: Atmospheres*, *127*, e2021JD036146. https://doi.org/10.1029/2021JD036146

Received 4 NOV 2021
Accepted 16 JUN 2022

Author Contributions:
Conceptualization: J.-P. Montillet, G. Kermarrec, R. Sikonja, M. Haberreiter, W. Schmutz, T. Dudok de Wit
Data curation: J.-P. Montillet, W. Finsterle
Formal analysis: J.-P. Montillet, G. Kermarrec, M. Haberreiter, W. Schmutz, T. Dudok de Wit
Funding acquisition: J.-P. Montillet, W. Finsterle, W. Schmutz
Investigation: J.-P. Montillet, G. Kermarrec
Methodology: J.-P. Montillet, G. Kermarrec, T. Dudok de Wit
Project Administration: W. Finsterle, W. Schmutz

© 2022. American Geophysical Union. All Rights Reserved.

1. Introduction

Monitoring the Earth's radiation budget is a key aspect in understanding the anthropogenic contribution to climate forcing (Kren, 2015). Total solar irradiance (TSI) 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 (Egorova et al., 2018; Rind et al., 2014; Schmutz, 2021; Shapiro et al., 2017), 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) significantly bias 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. Satellite measurements show that TSI varies on all timescales with a pronounced quasi-periodicity signature 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, which dominate each timescale are complex and still under investigation within the solar physics community (Xiang, 2019; Yeo et al., 2017). Several studies (Fontenla et al., 2009; Kopp & Lean, 2011; Yeo et al., 2021) have shown that TSI variations on timescales of hours are a combination of sunspot blocking and an intensification due to bright faculae, plages and other elements. This makes forecasting and modeling the solar cycle more difficult. As all satellite observations are limited in time, constructing composites is a key aspect to the investigation of TSI 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 (Fröhlich & Lean, 2004; Mekaoui & Dewitte, 2008;

Software: J.-P. Montillet, G. Kermarrec, R. Sikonja, T. Dudok de Wit
Validation: J.-P. Montillet, G. Kermarrec, R. Sikonja, T. Dudok de Wit
Visualization: J.-P. Montillet
Writing – original draft: J.-P. Montillet, G. Kermarrec, M. Haberreiter, T. Dudok de Wit
Writing – review & editing: J.-P. Montillet

Wilson, 1997) produced TSI composite time series by daisy chaining all the available TSI observations, but 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. An approach along these lines was also employed by (Schöll et al., 2016) and (Haberreiter et al., 2017). The authors made use of all available solar spectral irradiance datasets.

Here, we present a different statistical approach, which is based on three key steps. 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 wavelet filtering to correct some unwanted correlations in the fused observations (i.e., bandwidth noise). The robustness of our approach is guaranteed via careful modeling of the TSI observations during the data fusion process. Various assumptions formulated by data scientists, can introduce biases in the data analysis. Some algorithms (Fröhlich & Lean, 2004; Mekaoui & Dewitte, 2008; Wilson, 1997) 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 merges datasets from subsequent solar missions 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 toward preconceived ideas of how the TSI should vary (Dudok de Wit et al., 2017).

The motivation to develop our new methodology is to advance the data-driven approach first adopted by Dudok de Wit et al. (2017). Our resulting TSI composite corresponds with the instrument-driven approach by the late Dr. Claus Fröhlich (Fröhlich, 2006) when calibrated in a similar way. We use his composite (Fröhlich, 2006) as a baseline for other aspects as well (i.e., Step 3 of the algorithm). However, the TSI community consensus composite developed by Dudok de Wit et al. (2017), is the reference time series against which our new results are compared.

Finally, recent studies (Dudok de Wit & Kopp, 2020; Scafetta et al., 2020; Schmutz, 2021) have debated about the existence of a trend in the TSI 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 peak amplitude of solar cycles 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. One possibility could be that the brightness of the quiet Sun shows a trend, as suggested by Shapiro et al. (2011). 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 Datasets

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 the distance from the sun (normalized to 1 AU), radial velocity of 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 operate simultaneously, although some days are devoid of observations. Note that *PMODv21a* is the new VIRGO/SOHO data set 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/ERBS and HF/NIMBUS-7 ERB datasets 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 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/SORCE). Earlier radiometers such as ERBE/ERBS observes the sun once

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Overview of the Datasets Used in This Study Including the Start and End Dates for Each Mission and the Latest Version Released by the Various Centers

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	19	02/2003	02/2020
TIM/TCTE	4	12/2013	05/2019
TIM/TSIS	3	11/01/2018	Active

every 14 days for 3 min on average, so that the stochastic noise properties of such sensors are different to those with a higher recording rate. Note that *active* in Table 1 means that the instrument is still operating. The data set from these missions ends in March 2021 for this study.

3. The 3-Step Method to Produce the 41-Year Long TSI Composite

3.1. Step 1: Merging Multiple Datasets 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 $\{i = [1, n]\}$), which are recorded using three different instruments. The noise for each observation is additive and uncorrelated between the instruments. The model of the observations is defined such as:

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

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) modeling the noise properties intrinsic to each instrument. The data fusion algorithm aims to merge the observations available at each epoch t_i in order to get a reliable estimate of the true signal s, that is, the solar activity (Feynman, 1982). We formulate the following assumptions: (a) the solar cycle is an unknown process (i.e., not a perfect sinusoidal signal with a 11.5 years 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

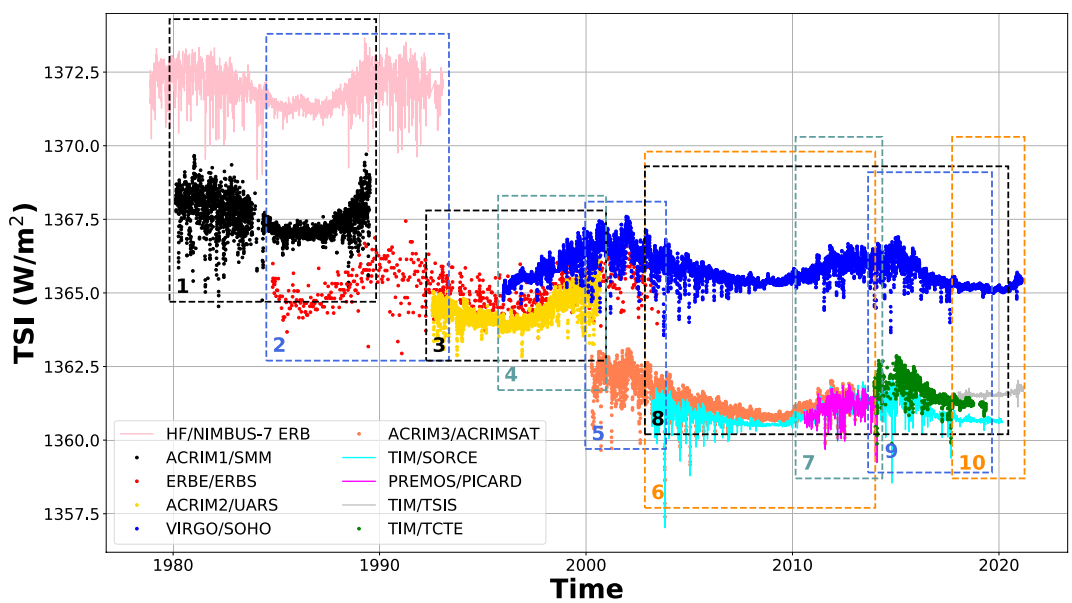


Figure 1. Various satellite missions which have conduscted total solar irradiance (TSI) observations since the late 1970s. We perform the fusion using the observations included in each box (dash lines).

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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); (b) we consider the noise on the measurements zero-mean Gaussian distributed. We can then estimate the parameters of the model of $s(t)$ via a maximum likelihood estimator (MLE). Therefore, we have $\mathbf{s} \sim GP\left(0, k_\theta\left(t_i, t_i\right)_{\{i=[1,n]\}}\right)$, 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-marginal likelihood log $p(\mathbf{y}|\mathbf{x})$, where $\mathbf{y}$ and $\mathbf{x}$ are the corrected observations, that is, $\mathbf{y} = [a(t_i), b(t_i), c(t_i)]$, and the corresponding time, that is, $\mathbf{x} = [t_i, t_i]$. 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, which has generated large datasets. To overcome this limitation, we approximate the exact GPs by utilizing sparse Gaussian processes, yielding a maximization problem of the lower bound of log $p(\mathbf{y}|\mathbf{x})$ following Bauer et al. (2016):

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

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 $\left(\left[\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_a^2, \sigma_b^2, \sigma_c^2, \ldots\right]\right)$. Next, we estimate the kernel k_θ by maximizing the right-hand-side of Equation 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). We also define comprehensively what is a GP and its application to TSI time series (including the Bayesian framework) in Appendix D. 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 Equation 2. A large number of points is necessary to avoid completely smoothing the short-term and long-term correlations due to the difference in the recording rate of the instruments. The computational complexity is on 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 $m = 3{,}000$). Now, the number of inducing points vary due to the size of the input datasets (i.e., size of the boxes defined in Figure 1). Appendix C further documents the influence of this parameter on the quality of the fused time series. We chose 2000 points, which provides 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 2,500 points or more if necessary. The shortest time series is when fusing PREMOS/PICARD, VIRGO/SOHO and TIM/SORCE (box 7), where we have ∼1,400 observations. In this case, we use 1,300 points. Note that the number of inducing points for the fusion of TIM/TSIS and VIRGO/SOHO is also constrained for the same reasons.

As described in Section 2, each instrument records the data with a different sampling rate. The fusion requires regularly sampled records with no gaps. We first regrid all the datasets with a sampling rate of 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/NIMBUS-7 ERB and ACRIM1/SMM which is February 1980.

Nonetheless, events resulting from short-term variations in solar activity lasting less than a few days 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 generally results in keeping only major events or underlying long-term solar events recorded by all the instruments at a specific time (t_i).

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

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 $\left(\alpha_q^2\right)$. We use a modified adaptive algorithm (Haykin, 2004) to daisy-chain all the sub-time series and build the 41-year composite as follows:

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

with t_i the time spanning the period 1978–2021 and daily sampling. The two overlapping time series 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 $\bar{\alpha_1^2} \le \bar{\alpha_2^2}$. This condition is necessary to guarantee that w is in the interval [0, 1]. We define w using the denominator $\alpha_2^2(t_i) + \alpha_1^2(t_i)$ in order to avoid any divergence. We exclude the case for which $\alpha_1^2(t_i) = \alpha_2^2(t_i) = 0$.

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 defined by Prša et al. (2016), which was derived as the averaged TSI value over Solar Cycle 23. This approach is also applied here, that is, 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 results in a filtering of the high frequency area of the spectrum of the composite time series. This is similar to the effect of a low pass filter, which would let the low frequencies pass through. 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: (a) it may mask some peaks at high frequencies linked with daily components; (b) the solar minimum can be affected in the time domain by the presence of long-term correlations. Unfortunately, this effect is theoretically unpredictable. Here we propose to make use of a wavelet filter to generate our final product. We intend to reconstruct empirically the high frequencies of our fused composite time series with respect to a reference one. We have chosen the one released by Fröhlich (2006), based on its properties in the time domain, discussed in the next section. To that end, the wavelet variance (WV) provides a robust mathematical framework to perform the rescaling of the high frequency noise of the composite time series (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. We make use of the maximum overlap discrete wavelet transform (MODWT), which has some advantages over the usual discrete wavelet transform: it avoids a downsampling process, unfavorable in some analyses(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 the Appendix B). For each of the 8 levels (or wavelet bands), one WV is estimated. We perform an analysis of WV versus scale in a log-log diagram following Abry and Veitch (1998) to identify the bandwidth noise at high

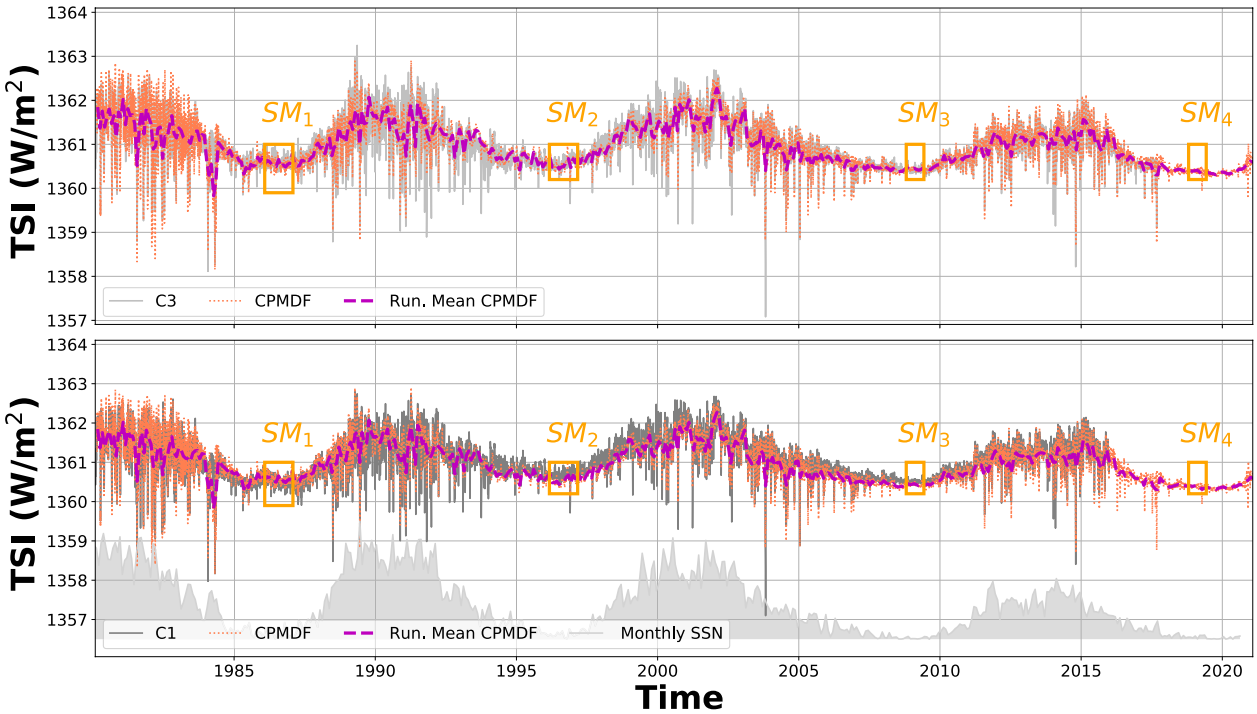


Figure 2. New composite *Composite PMOD- Data Fusion* (*CPMDF*, orange) based on merging 41 years of total solar irradiance (TSI) measurements. For comparison, *C3* (Fröhlich, 2006) and *C1* (Dudok de Wit et al., 2017) are also shown (gray line). A 30-day running mean of *CPMDF* is shown as a yellow/purple dashed 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.

frequency. In our case, this corresponds to the three first levels of decomposition, which are mostly affected by the data fusion process. The same WV decomposition is applied to the reference time series. The coefficients of the input time series are then rescaled with the WV ratio between the wavelet coefficients of the two time series. No assumption has to be made regarding the noise structure, for example, it may be a power-law or quantization noise. Intuitively, this is similar to 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, that is, the inverse MODWT. A comprehensive description of the wavelet filter is given in Appendix B.

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. Note that we do not include the ACRIM composite for the sake of clarity in this study. The ACRIM composite (Wilson, 1997) was shown to differ substantially with *C1*, *C2* and *C3*, as discussed in Dudok de Wit et al. (2017). Thus, it should also differ with our new composite and lead to the same conclusions. The new TSI composite is named *Composite PMOD- Data Fusion (CPMDF)*. In Appendix B, we discuss the composite without applying the wavelet filter. Figure 2 displays the composite time series overlaying *C1* and *C3*. The estimation of TSI at the solar minimum across the various solar cycles from 1980 to the present is estimated in Table 2.

Note that the solar minima are underlined in Figure 2 (see *SM* - 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. We then average the irradiance values over a 1-year interval centered on that date to produce the corresponding solar minimum value. The associated uncertainty is the variance of the measurements over the same period. When comparing the mean difference between the product and our time series

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Estimation of Total Solar Irradiance (TSI) at Solar Minimum (Minimum) Over the Last 41 Years From 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)

		Composite name							
		C1		C2		C3		CPMDF	
TSI level ($\mu \pm \sigma$ [W/m^2])		μ	σ	M	σ	μ	σ	μ	σ
Solar Cycle 21/22	*Minimum* (SM_1)	1360.51	0.13	1360.39	0.13	1360.59	0.12	1360.56	0.13
	$\Delta I_{21/22-20/21}$	–	–	–	–	–	–	–	–
Solar Cycle 22/23	*Minimum* (SM_2)	1360.69	0.14	1360.46	0.16	1360.57	0.15	1360.55	0.13
	$\Delta I_{22/23-21/22}$	**0.18**	**0.25**	**0.07**	**0.28**	**−0.02**	**0.27**	**−0.01**	**0.26**
Solar Cycle 23/24	*Minimum* (SM_3)	1360.53	0.04	1360.44	0.04	1360.42	0.06	1360.46	0.04
	$\Delta I_{23/24-22/23}$	**−0.16**	**0.18**	**−0.02**	**0.18**	**−0.15**	**0.21**	**−0.09**	**0.17**
Solar Cycle 24/25	*Minimum* (SM_4)	–	–	1360.43	0.07	–	–	1360.39	0.07
	$\Delta I_{24/25-23/24}$	–	–	**−0.01**	**0.11**	–	–	**−0.07**	**0.11**

Note. The new TSI composite is abbreviated to (*CPMDF*). The difference in 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).

over the various solar minima, the new composites agree with *C1* at 0.09 ± 0.04 W/m^2, *C2* at 0.08 ± 0.06 W/m^2, and *C3* at 0.03 ± 0.01 W/m^2. The difference is marginal for *C1*, *C2* and *C3*. Note that *C2* is rescaled to the nominal TSI value of 1361 W/m^2 adopted by the IAU 2015, averaged over Solar cycle 23. By applying this process, the calculated offset is equal to 2.44 W/m^2, which is due to the calibration based on the absolute level estimated from DIARAD/SOVIM. *C1* and *C3* are near the nominal TSI value (below 0.1 W/m^2) averaged over Solar cycle 23, due to their intrinsic processing.

Now, we analyze the Power Spectrum Density (PSD) of the composite time series *CPMDF* displayed in Figure 3. We can underline the four frequencies (11.5 years, 27, 9 and 7 days) related to solar activity and described by 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 into three areas (i.e., box *A*, *B* and *C*). The definition of the three areas follows the description of the photospheric activity. The latter associated with granulation, super-granulation and meso-granulation (Andersen et al., 1994; Fröhlich et al., 1997) generates fluctuations in TSI at different timescales. Due to our daily resolution, frequencies associated with phenomena lasting a few hours or less (i.e., granulation - (Fröhlich et al., 1997)) cannot be observed. Note that the instrumental effects due to the recording of the observations cannot be ignored, but it is difficult to decorrelate it from the other noise sources.

Box *A* shows a flattening of the curve at high frequencies. The PSD of the products *C1* and *C3* experience the same flattening at high frequencies. For comparison, 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) at a 1-min sampling rate. We can observe that the flattening effect 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) and Fröhlich et al. (1997) also show that the solar noise flattens in this frequency band. We can then conclude that the flattening curve of the PSD is due to the low-sampling rate in the TSI composites.

Box *B* is the power-law or the frequency ramp (between 0.06 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 could generate it. For example, it could be an unknown diffusion process associated with the sun's activity which could be modeled with a specific colored 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 colored noise:

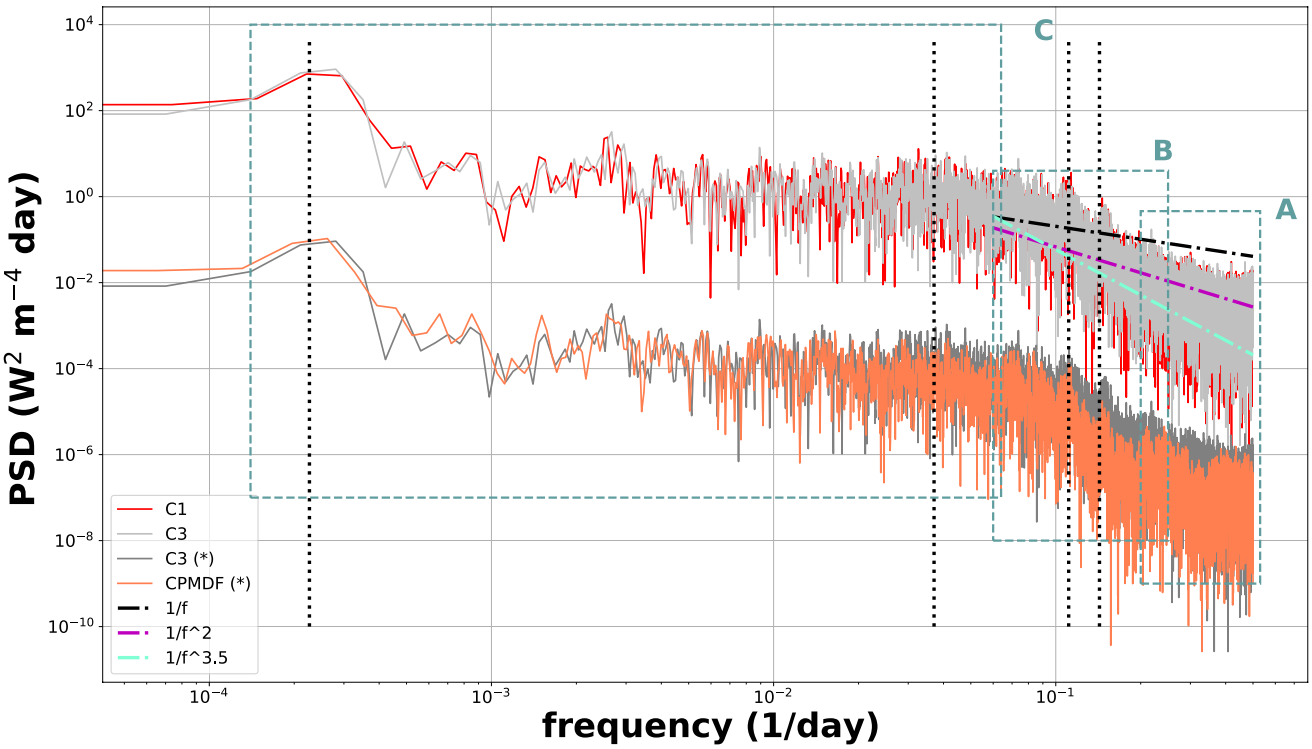


Figure 3. Power Spectrum Density (PSD) of total solar irradiance (TSI) *C1* (Dudok de Wit et al., 2017), *C3* (Fröhlich, 2006), together with the new TSI composite produced with the current method *Composite PMOD- Data Fusion* (*CPMDF*). The (*) means that the time series are shifted by rescaling the 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 (∼3 days) showing the flattening of the PSD; *B* is the power-law which is mainly due to colored noise (correlations between 20 and 6 days) within the time series; *C* emphasizes the low frequency associated with the stochastic and deterministic parts of the solar cycle and long-term correlations. The dashed lines are the various power-law models when varying the exponent, which are only shown for context. The vertical doted lines (black) mark the frequencies at 11.5 years, 27, 9 and 7 days (left to right).

flicker noise corresponds to β equal to 1, a random walk to β equal to 2, and white noise with β equal to 0 (J.-P. Montillet et al., 2021).

Moreover, in *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 subtracted various TSI time series from different missions (i.e., ACRIM1/SMM, ACRIM2/UARS, ACRIM3/ACRIMSAT, TIM/SORCE) in order to eliminate the solar cycle, resulting with the stochastic properties only, that is, a mix of noises between the two instruments. Figure 4 displays the PSD of the difference of the 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, mainly associated with photospheric activity. This result supports the conclusions in Dudok de Wit et al. (2017). Note that we cannot exclude the short/long-term correlations due to instrumental effects.

Finally, box *C* is associated with the low frequencies (0.06–0.00015 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), including perhaps also some long-term instrumental artifacts. In the appendices, Figure Appendix C.3 shows the spectrum of *C1* with and without the solar cycle. To remove the general trend of this cycle, we subtract the time series using a running mean with a 5-day window. We clearly see that the low frequencies in box *C* have the lowest power in the PSD after subtracting the running mean, hence supporting our assumption. In addition, this frequency band also contains some of the colored 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.

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

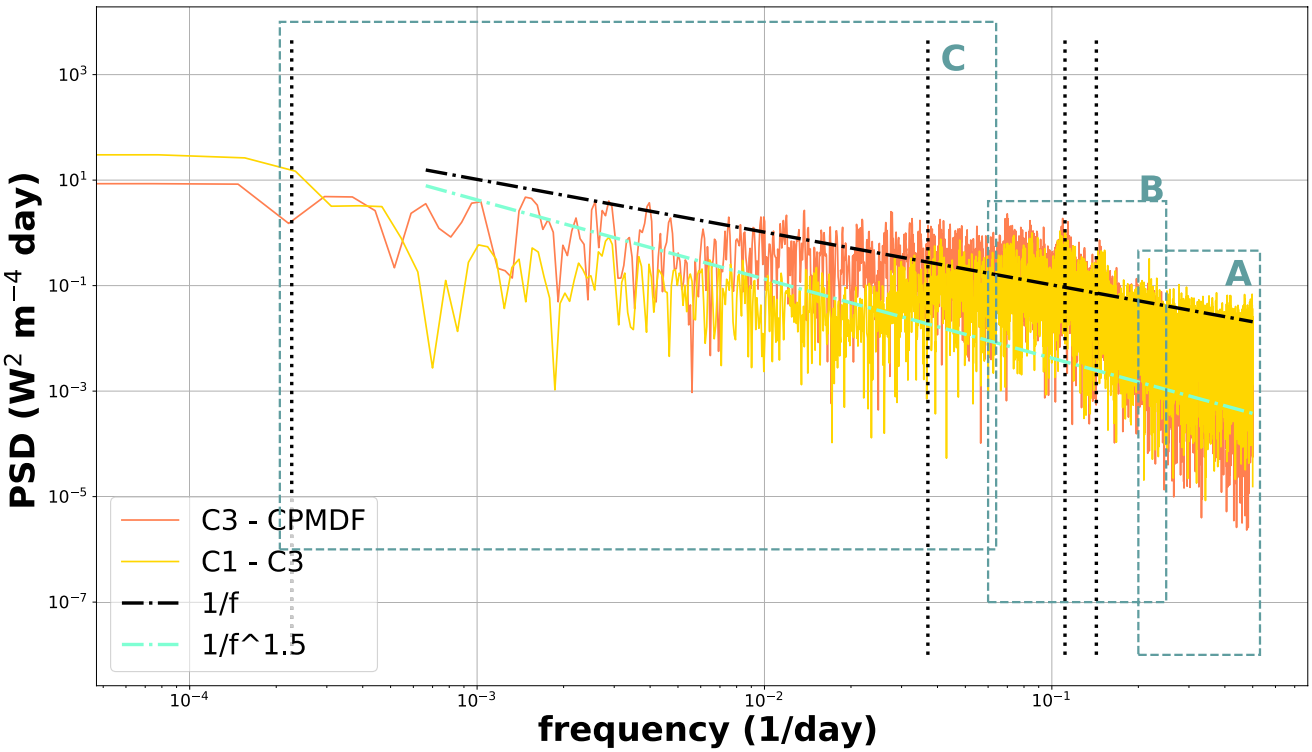


Figure 4. Power Spectrum Density of the difference of composite time series including *C3 - Composite PMOD- Data Fusion* (*CPMDF*), *C1 -C3*. The vertical doted lines (black) mark the frequencies at 11.5 years, 27, 9 and 7 days (left to right).

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: (a) the variations between consecutive solar minima, and (b) the global fluctuations (general trend) over the duration of the time series.

The estimation of the variations between two consecutive solar minima is challenging based on the PSD analysis of the composite time series. In order to be statistically robust, one needs to take into account the long-term correlations generated by the colored 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., $\Delta I_{22/23-21/22}$). Overall, the fluctuations of the 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 it is negative for the other composites. This disagreement could be due to the processing of the TSI observations for the first missions (e.g., HF/NIMBUS-7 ERB, ERBE/ERBS) discussed in Section 3.1. The fluctuation between the other solar cycles (i.e., Solar Cycle 22/23, 23/24, 24/25) is more homogeneous in terms of the sign value (i.e., negative for all of them). The averaged value is -0.10 ± 0.37 W/m^2 and -0.04 ± 0.15 W/m^2 for $\Delta I_{23/24-22/23}$ and $\Delta I_{24/25-23/24}$, respectively. However, any trend or large-scale fluctuation is downplayed by the large uncertainties associated with the difference between solar minima - up to 4 times the value.

For the study of the global trend (large-scale fluctuation) over the entire TSI composite time series, 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 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

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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), Dewitte and Nevens (2016) (C2), Fröhlich (2006) (C3) and by Applying (i.e., CPMDF) the Wavelet Filter

TSI level ($\mu \pm \sigma$ [W/(m^{2}yr)])				
Period	1980–2021		1987–2021	
Amplitude solar trend	μ	σ	μ	σ
C1	−0.004	0.006	−0.001	0.009
C2	−0.001	0.006	−0.003	0.007
C3	−0.009	0.006	−0.011	0.009
CPMDF	−0.014	0.004	−0.010	0.006

Note. CPMDF, *Composite PMOD- Data Fusion*; GGM, General Gauss-Markov model; MLE, maximum likelihood estimator; TSI, total solar irradiance. Two different time periods are chosen. *yr* means year.

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 into two time intervals 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 solar cycle 21 allows the trend between the difference in solar minima to be debated.

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 product *CPMDF* 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 gives -0.005 ± 0.003 W/(m^{2}yr) and -0.005 ± 0.004 W/(m^{2}yr) for the periods 1980–2021 and 1987–2021, respectively. When we include the new product, the average trend is -0.004 ± 0.003 W/(m^{2}yr) and -0.004 ± 0.004 W/(m^{2}yr) for the same periods. The overall estimate, using the results from all the TSI composite time series and both periods, is equal to $\sim -0.004 \pm 0.004$ 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 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 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

We have merged 41 years of satellite observations using data fusion in order to produce a TSI composite time series, 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, including the TSI community consensus time series *C1*. The results show that the mean value of the difference over the solar minima is below the 1 sigma confidence interval of 0.2 W/m^2, that is, a maximum of 0.09 W/m^2 for *C1* and a minimum of 0.03 ± 0.01 W/m^2 for *C3*. In terms of comparing the frequency spectrum, 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 the meso-granulation, granulation and p-modes frequency bands.

Second, the power spectrum experiences a power-law between 4 and 20 days which could correspond to an unknown diffusion process. When removing the solar cycle by differencing two TSI composite time series, a frequency ramp (power-law noise) is observed over the whole frequency band. The power-law exponent varies within [1, 2]. It highlights the presence of long-term correlations from solar noise and perhaps instrumental 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 $\sim -0.004 \pm 0.004$ 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 reach any conclusions about the

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

existence of a linear trend in the TSI composite time series. Any visual effects or short-term trends are most likely related to the colored 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).

Appendix A: Model Descriptions for Estimation of the Linear Trend

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

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

where $E\{.\}$ is the expectation operator. The vector $\psi = [\psi(t_1), \psi(t_2), \ldots, \psi(t_L)]$ 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 colored 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 of the type of colored 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).

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

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 ($2\pi fq_j$) and fq_j are different frequencies (e.g., 11.5 years, 27, 9 and 7 days) which are determined by analyzing 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 MLE. To recall Bos et al. (2020), the log-likelihood for a time series of length n can be rewritten as:

$$\ln(Lo) = -\frac{1}{2}\left[n\ln(2\pi) + \ln(\det(\mathbf{C})) + (\mathbf{x_0} - \mathbf{Az})^T\mathbf{C}^{-1}(\mathbf{x_0} - \mathbf{Az})\right] \tag{A3}$$

This function must be maximized. Assuming that the covariance matrix $\mathbf{C}$ is known, then it is a constant and does not influence finding the maximum. $\mathbf{C}$ is equal to $E\{\psi^T\psi\}$ as defined by Equation A1. The term $(\mathbf{x_0} - \mathbf{Az})$ represents the TSI observations minus the fitted model. Note that $(\mathbf{Az})$ is the matrix notation of s_0. The last term can be written as $\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}$ and it is a quadratic function, weighted by the inverse of matrix $\mathbf{C}$. 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 (~15,330 observations).

The particularity of the TSI composite frequency spectrum has been discussed above, and in particular its flattening at high frequencies. Therefore, a simple power-law noise model as described in Section 3 with the covariance $\mathbf{J}(\beta)$ is not appropriate for our ML estimation. Instead, we use the General Gauss-Markov model (GGM) noise model which has the advantage of flattening at high frequencies. The power spectrum density (PSD) of the GGM noise is defined by Bos et al. (2020) 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 following the recommendation of He et al. (2019). Also from Equation A4, if ϕ is set equal to 1 then 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.

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix B: Description of the Wavelet Filter

This section comprehensively describes the wavelet filter proposed in Section 3.3. We discuss the steps to correct some of the long-range dependency introduced by the data fusion process to generate the final TSI *CPMDF*. In this section, let us call *CPMDFa*, the composite time series before applying the wavelet filter.

B1. Methodology

The method for rescaling the high-frequency power of CPMDF can be divided into 4 steps:

1. The MODWT decomposition : discrete versus maximum overlap discrete wavelet transform

The first step of the wavelet filter is the wavelet decomposition of *CPMDFa* itself. To that end, the time series is usually broken down into a scaled and shifted version of a chosen mother wavelet. Because the sample size of the orthogonal discrete wavelet transform (DWT) is limited to a power of 2, the number of scaling and wavelet coefficients at each level of resolution decreases by the same factor. Unfortunately, this results (i) in a loss of information as the level of decomposition increases, as well as (ii) in the introduction of ambiguities in the time domain. These effects are unwanted when performing a scale-by-scale variance analysis. We circumvent the drawback of the DWT by using the maximal overlap discrete wavelet transform (MODWT, (Cornish et al., 2006)), which carries out the same steps as the DWT without a sub-sampling process. Mathematically, the MODWT is a convolution operation that can be formulated as circular filter operations of the original time series using two quadrature mirror filters, 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. The MODWT is known as a shift-invariant wavelet transform. It is a highly redundant version of the DWT and is considered ideal for time-series analysis, as it accommodates any sample size. In this contribution, we use the MATLAB wavelet package from Mathworks (https://ch.mathworks.com/).

Whereas low scales of the MODWT decomposition are related to long periodic behavior, high scales focus on brief phenomena. The proposed wavelet filter aims to rescale the power of the high frequencies of *CPMDFa*, which are filtered out by the data fusion, and recompose the obtained time series: the choice of the wavelets should take this application into consideration. Here, we follow the work of Cornish et al. (2006) who proposed to use the least asymmetric (LA) wavelets. These wavelets exhibit near symmetry about the filter midpoint which allows a good alignment of the reconstruction with the original time series by circularly sifting the coefficients. More specifically, we use the LA(4): this wavelet has a nearly linear phase response and is optimal for reconstruction. Compared to the Haar wavelet, it has less leakage, which makes it more appropriate to rescale the variance of the chosen levels of decomposition (see step 2). Here the band-pass filtering is more accurate and allows a better control over the rescaled levels of decomposition at high frequencies. We further note that the 4 vanishing moments produce wavelet coefficients vectors that are nearly stationary; this is favorable to analyze the variance. We then chose to decompose the signal into 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.

2. The wavelet variance (WV) decomposition

In a second step, we compute the variance of each decomposition level time series, on a scale-by-scale basis. The variances are called the wavelet variances (WV). The WV can be interpreted as the variance of a process after filtering by a wavelet bandpass filters (Percival & Guttorp, 1994). Our method to rescale the WV is based on an unbiased wavelet estimator developed by McCoy and Walden (1996) and Abry and Veitch (1998) for correlation analysis. The WV multiscale analysis provides an efficient and highly robust estimator of the fractal parameter of a process. We recall that fractal processes have a power of the form $1/f^{\beta}$ for a range of frequency f close to 0. It can be shown that the WV versus its scales have a logarithmic linear relationship for such processes: 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 at high frequencies induced by the windowing used to fuse TSI from different datasets.

3. Rescaling the WV

The procedure is inspired by 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 that under consideration. When dealing with a mix of unknown bandwidth fractal noises, the scales at which a given noise is present have to be identified in a first step in the $\log_2$-diagram WV versus scales. As an example, a white noise exhibits a slope of -1, flicker

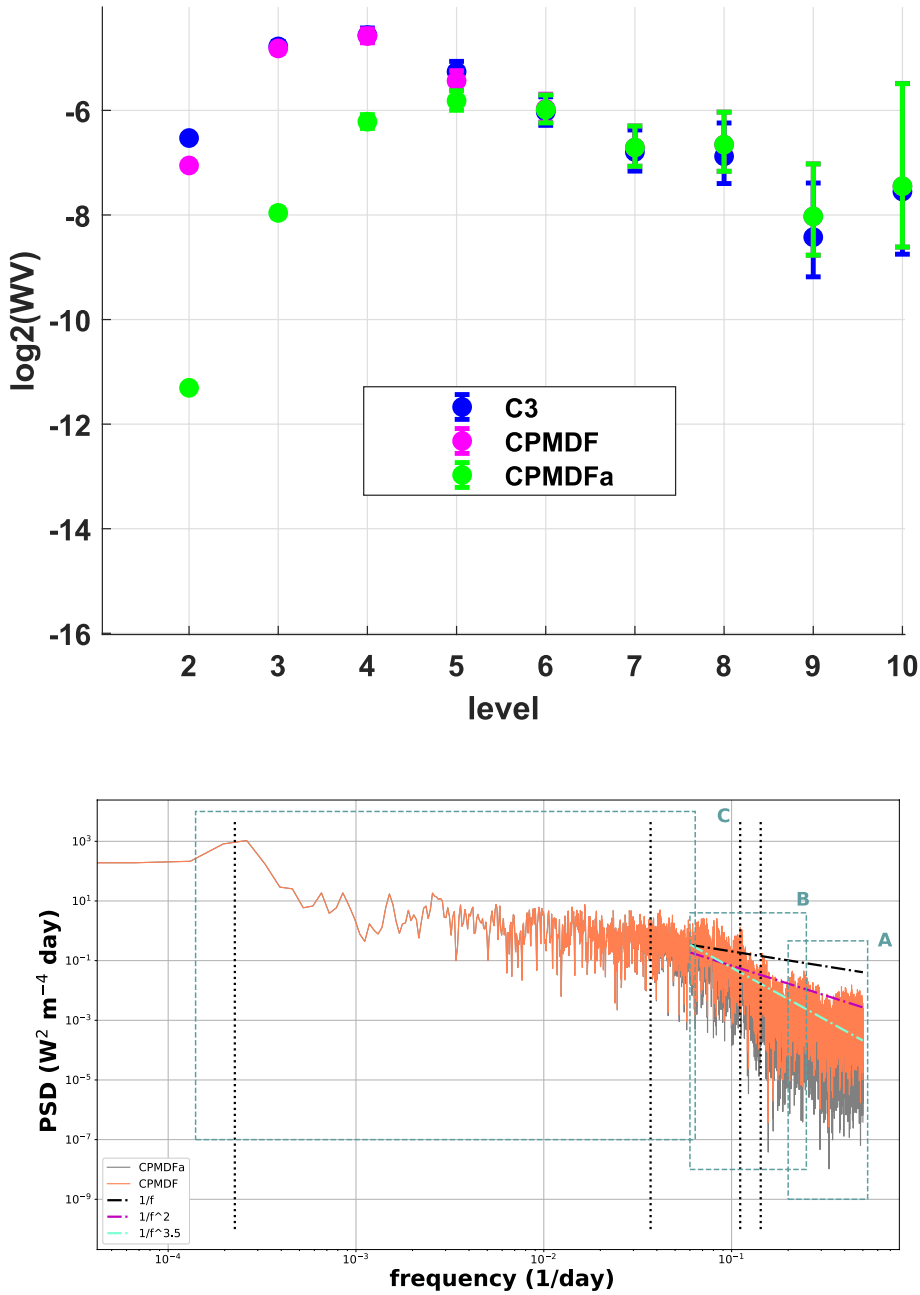


Figure B1. (top) Wavelet Variance (WV) decomposition of *C3* (blue), *CPMDFa* (green) and the filtered time series (pink). The X-axis is the level of decomposition, whereas the Y-axis is the WV value on a log scale. (bottom) Effect of applying the wavelet filter in the frequency domain: without *CPMDFa* (gray) and with *CPMDF* (orange).

noise has a slope of 0 and, a random-walk has a slope of 1. 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 or rescaling procedure using *C3*. Figure B1.A is a $\log_2$-diagram showing the WV versus scales. Our reference time series is marked with blue dots, together with the fused time series before (green) and after (pink) filtering. The WV of both time series lies on a straight line. For the reference time series, the slope approximately corresponds to a coefficient equal to -3 in the WV $\log_2$-diagram. The first WV contains most probably an additional WV component as it is slightly over the line drawn from scale 2 and 3. From scale 4, the WV spectrum clearly changes its shape, which corresponds to the Matérn process (saturation at low frequency). This visual analysis of the $\log_2$-diagram shows that the sought after high frequency correlated noise is present between the scales 1 and 3.

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Here we propose to rescale the WV of *CPMDFa* in order to fit the reference one (i.e., *C3*). The resulting WV spectrum is shown in the green dots in Figure B1.A. Intuitively, we have re-introduced high frequency noise to obtain the reference decay of the WV spectrum (or similarly that of the power spectral density).

4. Recomposition of the rescaled TSI

Finally, we recompose the filtered time series by using the inverse function inserting our new WV values. Note that we use the function called Inverse MODWT, abbreviated as IMODWT, which is 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 *CPMDFa*, well within the interval of confidence of the reconstruction (0.2 W/m^2) as discussed in Section 4. Note that the MODWT and IMODWT require $O(n.\log_2.n)$ multiplications (Percival & Walden, 2000).

B2. Discussion of the Effect of Applying the Wavelet Filter

It is worth showing the impact of applying the wavelet filter on the TSI time series composite. Figure B1.B displays the PSD of the time series before and after the filter. The difference can only be seen in Box *B*. The steepness of this ramp is more accentuated for *CPMDFa* than in *CPMDF*. The power-law is between 2.5 and 3 for the previous releases *C1*, *C2*, *C3* and *CPMDF*, but between 3 and 3.5 for *CPMDFa*. This increase steepness is a weakness of the data fusion process as discussed in Section 2, 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.

Appendix C: Additional Figures and Remarks

C1. Remarks on the Inducing Points

Figure C1 shows the variations in both time and frequency when fusing VIRGO/SOHO, TIM/SORCE and ACRIM3/ACRIMSAT (box 8 in Figure 1). Figure C2 displays the associated PSD. It is difficult to find an optimal number, because above 1500 points the time fluctuations do not show many differences (i.e., continuity of the spectrum, amplitude of the frequencies associated with the solar cycle).

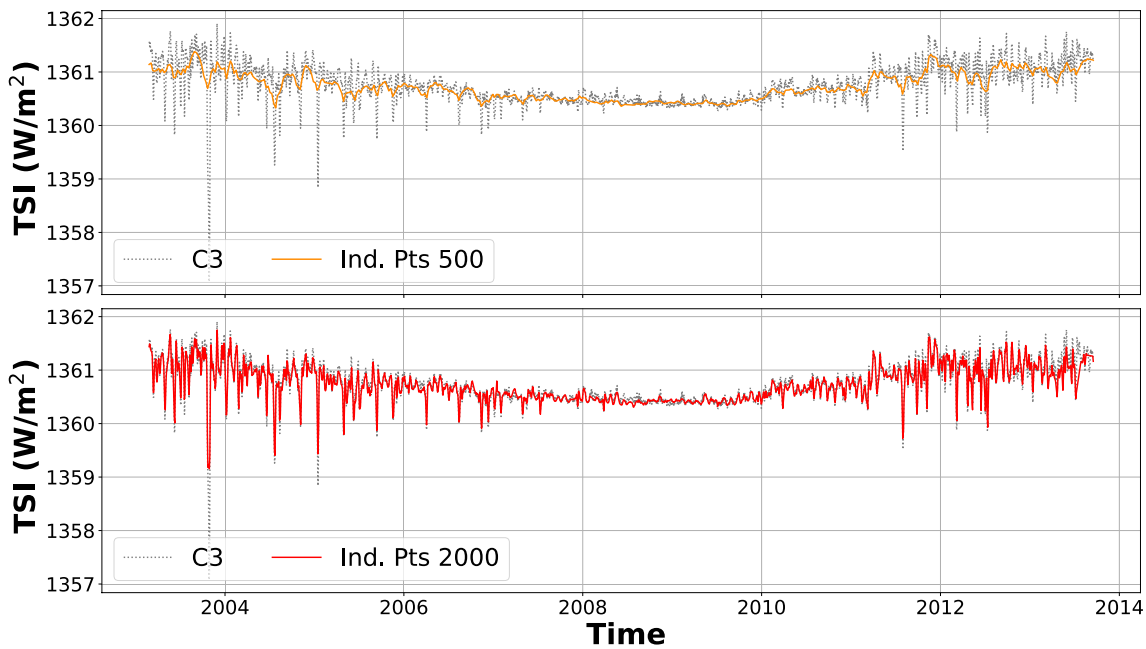


Figure C1. Time series of *C3* (Fröhlich, 2006) and the sub-time series in Box 8 (see Figure 1) fusing VIRGO/SOHO, ACRIM3/ACRIMSAT and TIM/SORCE using various numbers of inducing points (500, 2,000). The sub-time series are aligned on *C3*.

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

C2. Additional Figures

The following figures (C2, C3 and C4) illustrate the dicussions in Section 4.

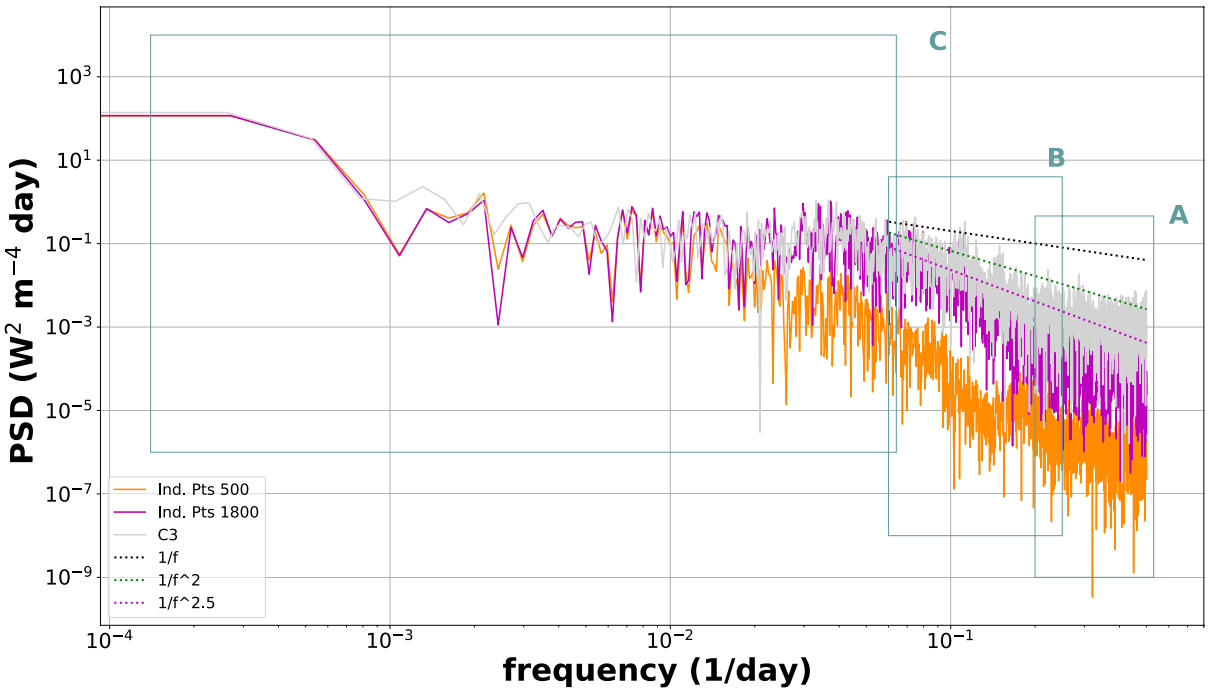


Figure C2. Power Spectrum Density (PSD) of total solar irradiance (TSI) *C3* (Fröhlich, 2006) and the sub-time series in Box 8 fusing VIRGO/SOHO, ACRIM3/ACRIMSAT and TIM/SORCE using various numbers of inducing points (500, 1,800). Box *A*, *B* and *C* refer to the different sections of the PSD: *A* is centered on the high frequency (~3 days), which shows flattening of the PSD; *B* is the power-law which is mainly due to colored noise (correlations between 20 and 6 days) within the time series; *C* emphasizes the low frequency associated with the stochastic and deterministic parts of the solar cycle and long-term correlations, as well as instrumental artifacts. The dashed line is the power-law model when varying the exponent (shown for context).

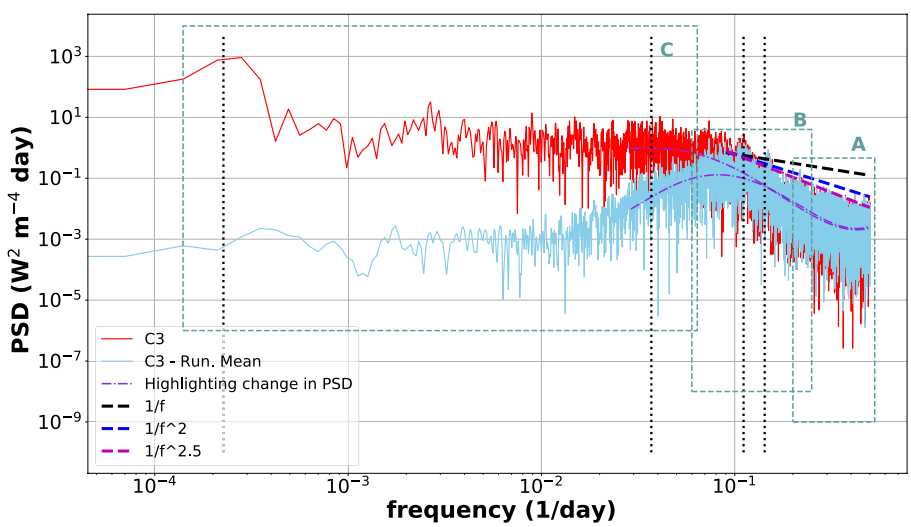


Figure C3. Power Spectrum Density (PSD) of total solar irradiance (TSI) *C3* (Fröhlich, 2006) with and without partially removing the solar cycle via a running mean (with a 5-day window). Boxes *A*, *B* and *C* refer to the different sections of the PSD. The dash-dotted lines are the various power-law 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-dotted lines highlight the change of power in box *C* before and after removing the solar cycle.

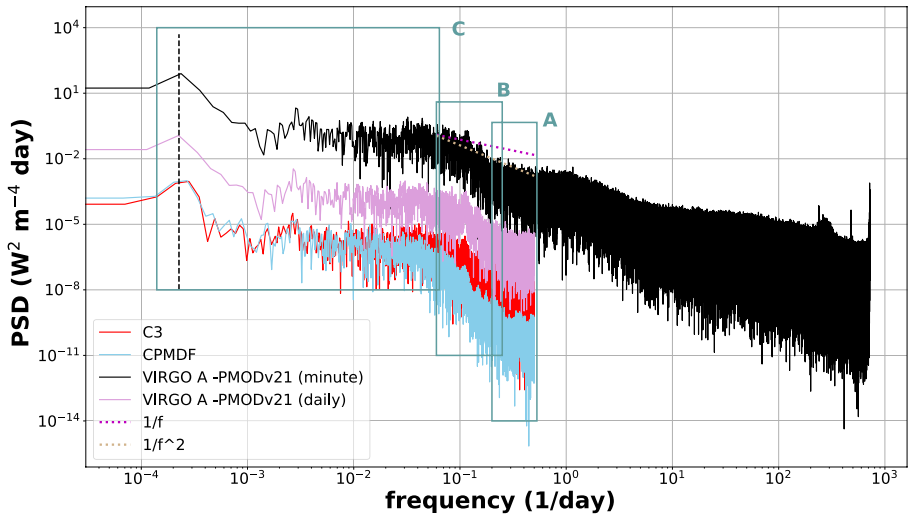


Figure C4. Power Spectrum Density (PSD) of various products: VIRGO/SOHO with degradation correction (VIRGO-A) from PMO6v21 with 1-min and daily sampling rate. *C3* is the total solar irradiance (TSI) composite produced by Fröhlich (2006). *Composite PMOD- Data Fusion* is our new TSI composite including the wavelet filter. Note that for clarity the different spectra have been rescaled by multipying them by −6, −6, −4 and −8.5 W^2 m^{-4} day following the order of the legends from the top. Boxes *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 (shown for context). The vertical dashed line emphasizes the 11.5 years peak.

Appendix D: Assumptions and Discussion About the Data Fusion Algorithm

To further clarify the assumptions in order to fuse the various observations in Step 1 in Section 3.1, we discuss in details how the Gaussian processes are applied in the specific case of the TSI data. We also define the Bayesian framework used to perform the fusion based on Kolar et al. (2020). The notations in Section 3.1 are used in this appendix.

Let us define the notion of *Gaussian process* as used to model the solar cycle and its variations at various scales. According to Rasmussen and Williams (2006), a Gaussian process is a generalization of the Gaussian probability distribution. A probability distribution describes random variables which are scalars or vectors (for multivariate distributions). One can loosely think of a function as a very long vector, each entry in the vector specifying the function value at a particular input (e.g., time). A Gaussian Process is a flexible distribution over functions, with many useful analytical properties. In other words, a Gaussian process is a finite linear combination of random variables with a multivariate normal distribution, completely defined by the first and second-order statistics—the mean μ and the covariance matrix - or kernel- $k_\theta(\mathbf{x}, \mathbf{x})$. The kernel can be defined by the addition or multiplication of other kernels (periodic, linear, white, Matérn …). We assume μ to be zero, without loss of generality. The covariance function determines properties of the functions, such as smoothness, amplitude, and so on.

Let us model the function of interest, that is, the solar cycle s in our TSI data set, using a GP prior, noisy observations and the associated time for each observation such as $\mathbf{s} \sim GP\left(0, k_\theta\left(t_i, t_i\right)_{\{i=[1,n]\}}\right)$. s is not a deterministic signal (i.e., perfect sinusoid with an 11-year cycle) and its variations are random (no a priori knowledge). In terms of probability distribution, we can state:

$$p(\mathbf{s}) = \mathcal{N}(\mathbf{s}; 0, k_\theta)$$

$$p(\mathbf{y}|\mathbf{s}) = \Pi_{i=1}^{n} \mathcal{N}\left(y_i; s_i, \sigma^2\right)$$

$p(\mathbf{s})$ is the probability distribution of s. $p(\mathbf{y}|\mathbf{s})$ is the conditional probability of s knowing $\mathbf{y} = [y_i]$. In Section 3.1, the observation y_i is not a scalar, but a vector due to the number of input TSI time series to fuse, that is, $\mathbf{y}_i = [a(t_i), b(t_i), c(t_i)]$ with an uncorrelated noise $\sigma^2\mathbf{I}$. The functions a, b and c are defined in the model of the observations in Equation 1. $\sigma^2\mathbf{I}$ is (assuming uncorrelated measurements) formulated as diag $\left(\left[\sigma_a^2, \sigma_b^2, \sigma_c^2\right]\right)$. Now, the covariance matrix of $\mathbf{y}$ is defined as $k_\theta(\mathbf{x}, \mathbf{x}) + \sigma^2\mathbf{I}$. $\mathbf{x} = [t_i, t_i]$ is the concatenation of times associated with each input time

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

series. All the hyperparameters defining the kernel are in the vector θ. The common approach is to estimate these hyperparameters to model the desired signal using a (marginal) MLE:

$$\theta^* = \underset{\theta}{argmax}\, p(\mathbf{y}|\theta)$$

and the estimation of s (s^*) via:

$$p(\mathbf{y}^*|\mathbf{y}) = \frac{p(\mathbf{y}^*, \mathbf{y})}{\mathbf{p}(\mathbf{y})} = \int p(\mathbf{y}^*|\mathbf{s}^*)\, p(\mathbf{s}^*|\mathbf{s})\, p(\mathbf{s}|\mathbf{y}) d\mathbf{s} d\mathbf{s}^*$$

Note that * means the estimated parameter or signal. While the marginal likelihood, the posterior and the predictive distribution all have closed-form Gaussian expressions, the cost of evaluating them scales as $O(n^3)$ due to the inversion of $k_\theta(\mathbf{x}, \mathbf{x}) + \sigma^2\mathbf{I}$, which is impractical for large datasets. To overcome this limitation, we approximate the kernel matrix $k_\theta(\mathbf{x}, \mathbf{x})$ with a low-rank matrix Q_θ described in Section 3.1 as $Q_\theta = k_\theta(\mathbf{x}, \mathbf{u})k_\theta(\mathbf{u},\mathbf{u})^{-1}k_\theta(\mathbf{u}, \mathbf{x})$. $\mathbf{u}$ is a vector of inducing points. It is important to underline that the vector of inducing points $\mathbf{u}$ introduced in this approximation is crucial. A proper training procedure (size of the vector of inducing points) permits us to learn about the stochastic properties of the data at various scales, which allows long-term and short-term correlations to be taken into account. Various methods exist to perform the MLE using the low-rank matrix with additional assumptions. Readers can refer to Bauer et al. (2016) for a comprehensive description on this topic. Here, we use the maximization of the lower bound derived from the log-marginal likelihood log $p(\mathbf{y}|\mathbf{x})$ based on the variational free energy method developed in Bauer et al. (2016) and discussed in Kolar et al. (2020). With this approximation, one needs to consider s as a sparse GP, a special class of GPs. With many data sample (i.e., $>10^3$), the error due to the approximating the GP as a sparse GP tends to 0 according to Kolar et al. (2020).

The first test of data fusion was carried out when fusing the data recorded by the radiometer PMO6V on board of the SOHO/VIRGO mission from the two channels VIRGO/PMO6V-A and VIRGO/PMO6V-B described in Finsterle et al. (2021). Kolar et al. (2020) optimzed the data fusion process using dedicated simulations. The latter were similar when more channels/observations were analyzed. We spare the patience of the Reader by describing redundant simulations.

Finally, we would like to comment further on the difference of definitions between the stochastic and solar noises described in Section 3. In our work, solar noise is defined to cover solar activity and the small-timescale processes which produce short-term variations. The stochastic noise takes into account both the instrumental effect and the small-timescale solar activity. That together produces the random variability (i.e., white noise) in time at a small scale. The instrumental noise defines all the short-term and small amplitude solar variations which cannot be properly observed by the radiometer due to its characteristics (observation rate, distance to the event, …). Statistically, it is an uncorrelated zero-mean Gaussian noise. On a large scale, the stochastic noise can be modeled as having a power law dependency with respect to frequency, which could be related to diffusive effects. We are here speaking about the long-term correlations. It is related to the solar noise with the diffusion effects within the solar cycle, the sun's rotation period, … and so on.

Data Availability Statement

The new composite *CPMDF* can be obtained from the open archive repository www.astromat.org (J. Montillet et al., 2022). It is also presented on www.pmodwrc.ch/en/?s=TSI+Composite (last accessed 07 June 2022) for additional information. The TSI composite *C1* is available for downloading at http://www.issibern.ch/teams/solarirradiance (last accessed 07 June 2022). The data related to the monthly/daily mean sunspot numbers are retrieved from http://www.sidc.be/silso/datafiles (last accessed 07 June 2022). *TIM/SORCE*, TCTE/TIM and *TIM/TSIS* time series are downloaded from https://lasp.colorado.edu/home/sorce/data/tsi-data (last accessed 07 June 2022). *PREMOS (v1)* can be accessed at http://idoc-picard.ias.u-psud.fr/sitools/client-user/Picard/project-index.html (last accessed 07 June 2022). *PMODv21a* is available at https://www.pmodwrc.ch/en/research-development/space/soho/ (last accessed 07 June 2022).

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Acknowledgments
We acknowledge the life-long dedication to the total solar irradiance (TSI) community of the late Dr. Claus Fröhlich in memoriam as a 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*). Dr. J.-P. Montillet, Dr. W. Finsterle, Dr. M. Haberreiter and Prof. W. Schmutz gratefully acknowledge the support from the 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. The authors thank the anonymous reviewers whose comments/suggestions helped improve and clarify this manuscript.

References

Abry, P., & Veitch, D. (1998). Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory*, *44*(1), 2–15. https://doi.org/10.1109/18.650984

Andersen, B. N., Leifsen, T., & Toutain, T. (1994). Solar noise simulations in irradiance. *Solar Physics*, *152*(1), 247–252. https://doi.org/10.1007/BF01473211

Bauer, M., van der Wilk, M., & Rasmussen, C. E. (2016). *Understanding probabilistic sparse Gaussian process approximations*. In *Nips* (Vol. 2016). 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. *Geophysical Journal International*, *96*(3), 1423–1430. https://doi.org/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). Springer International Publishing. https://doi.org/10.1007/978-3-030-21718-1_2

Cocchi, M. (2019). In *Data fusion methodology and applications* (Vol. 31). Elsevier. https://doi.org/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 Meteorology*, *119*(2), 339–374. https://doi.org/10.1007/s10546-005-9011-y

Davis, J., Wernicke, B., & Tamisiea, M. (2012). On seasonal signals in geodetic time series. *Journal of Geophysical Research*, *117*(B1), B01403. https://doi.org/10.1029/2011JB008690

Dewitte, S., & Nevens, S. (2016). The total solar irradiance climate data record. *The Astrophysical Journal*, *830*(1), 25. https://doi.org/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. *Geophysical Research Letters*, *44*(3), 1196–1203. https://doi.org/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 of Earth Science*, *6*, 206. https://doi.org/10.3389/feart.2018.00206

Feynman, J. (1982). Geomagnetic and solar wind cycles, 1900–1975. *Journal of Geophysical Research*, *87*(A8), 6153–6162. 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. https://doi.org/10.1038/s41598-021-87108-y

Fontenla, J., Curdt, W., Haberreiter, M., Harder, J., & Tian, H. (2009). Semiempirical models of the solar atmosphere. III. Set of non-LTE models for far-ultraviolet/extreme-ultraviolet irradiance computation. *The Astrophysical Journal*, *707*(1), 482–502. https://doi.org/10.1088/0004-637X/707/1/482

Fröhlich, C. (2006). Solar Irradiance Variability Since 1978. Revision of the PMOD Composite during Solar Cycle 21. *Space Science Reviews*, *125*(1–4), 53–65. https://doi.org/10.1007/s11214-006-9046-5

Fröhlich, C., Andersen, B., Appourchaux, T. e. a., Berthomieu, G., Crommelynck, D. A., Domingo, V., et al. (1997). First results from VIRGO, the experiment for helioseismology and solar irradiance monitoring on SOHO. *Solar Physics*, *170*, 1–25. https://doi.org/10.1023/A:1004969622753

Fröhlich, C., & Lean, J. (2004). Solar radiative output and its variability: Evidence and mechanisms. *Astronomy and Astrophysics Review*, *12*(4), 273–320. https://doi.org/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*(503), 1021–1030. https://doi.org/10.1080/01621459.2013.799920

Haberreiter, M., Schöll, M., Dudok de Wit, T., Kretzschmar, M., Misios, S., Tourpali, K., & Schmutz, W. (2017). A new observational solar irradiance composite. *Journal of Geophysical Research: Space Physics*, *122*(6), 5910–5930. https://doi.org/10.1002/2016JA023492

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

He, X., Bos, M., Montillet, J., & Fernandes, R. (2019). Investigation of the noise properties at low frequencies in long GNSS time series. *Journal of Geodesy*, *93*(9), 1271–1282. https://doi.org/10.1007/s00190-019-01244-y

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

Kopp, G. (2016). Solar variability magnitudes and timescales. *Space Science Reviews*, *6*, A30. https://doi.org/10.1051/swsc/2016025

Kopp, G., & Lean, J. L. (2011). A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, *38*(1), L01706. https://doi.org/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 & Graphical Statistics*, *5*(1), 26–56. https://doi.org/10.2307/1390751

Mekaoui, S., & Dewitte, S. (2008). Total solar irradiance measurement and modelling during cycle 23. *Solar Physics*, *247*(1), 203–216. https://doi.org/10.1007/s11207-007-9070-y

Montillet, J., Finsterle, W., Schmutz, W., Haberreiter, M., Dudok de Wit, T., Kermarrec, G., & Sikonja, R. (2022). Composite PMOD data fusion. [dataset]. Interdisciplinary Earth Data Alliance (IEDA). https://doi.org/10.26022/IEDA/112238, Accessed 2022-06-07.

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. *Nonlinear Processes in Geophysics*, *28*(1), 121–134. https://doi.org/10.5194/npg-28-121-2021

Percival, D., & Walden, A. (2000). *Wavelet methods for time series analysis*. Cambridge University Press. https://doi.org/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., et al. (2016). Nominal values for selected solar and planetary quantities: IAU 2015 Resolution B3. *The Astronomical Journal*, *152*(2), 41. https://doi.org/10.3847/0004-6256/152/2/41

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. The MIT Press - Massachusetts Institute of Technology.

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. https://doi.org/10.1175/JCLI-D-13-00136.1

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Scafetta, N., Wilson, R., Lee, J., & Wu, D. L. (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). In *Total solar irradiance measurements with PREMOS/PICARD* (Vol. 1531, pp. 624–627). American institute of physics conference series. https://doi.org/10.1063/1.4804847

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

Schöll, M., Dudok de Wit, T., Kretzschmar, M., & Haberreiter, M. (2016). Making of a solar spectral irradiance dataset I: Observations, uncertainties, and methods. *Journal of Space Weather and Space Climate*, *6*, A14. https://doi.org/10.1051/swsc/2016007

Shapiro, A., Schmutz, W., Rozanov, E., Schoell, M., Haberreiter, M., Shapiro, A. V., & Nyeki, S. (2011). A new approach to the long-term reconstruction of the solar irradiance leads to large historical solar forcing. *Astronomy and Astrophysics*, *529*, A67. https://doi.org/10.1051/0004-6361/201016173

Shapiro, A., Solanki, S., Krivova, N., Cameron, R., Yeo, K., & Schmutz, W. (2017). The nature of solar brightness variations. *Nature Astronomy*, *1*(9), 612–616. https://doi.org/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. 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. 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. *Atmospheric Chemistry and Physics*, *19*(3), 2079–2093. https://doi.org/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*, *109*(B3), B03412. https://doi.org/10.1029/2003JB002741

Wilson, R. (1997). Total solar irradiance trend during solar cycles 21 and 22. *Science*, *277*(5334), 1963–1965. https://doi.org/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*, 1–9. https://doi.org/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. https://doi.org/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. *Astronomy & Astrophysics*, *654*, A28. https://doi.org/10.1051/0004-6361/202141336

Yeo, K. L., Solanki, S. K., Norris, C. M., 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. https://doi.org/10.1103/physrevlett.119.091102

21698996, 2022, 13, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JD036146 by Portail BibCNRS INSU, Wiley Online Library on [31/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License