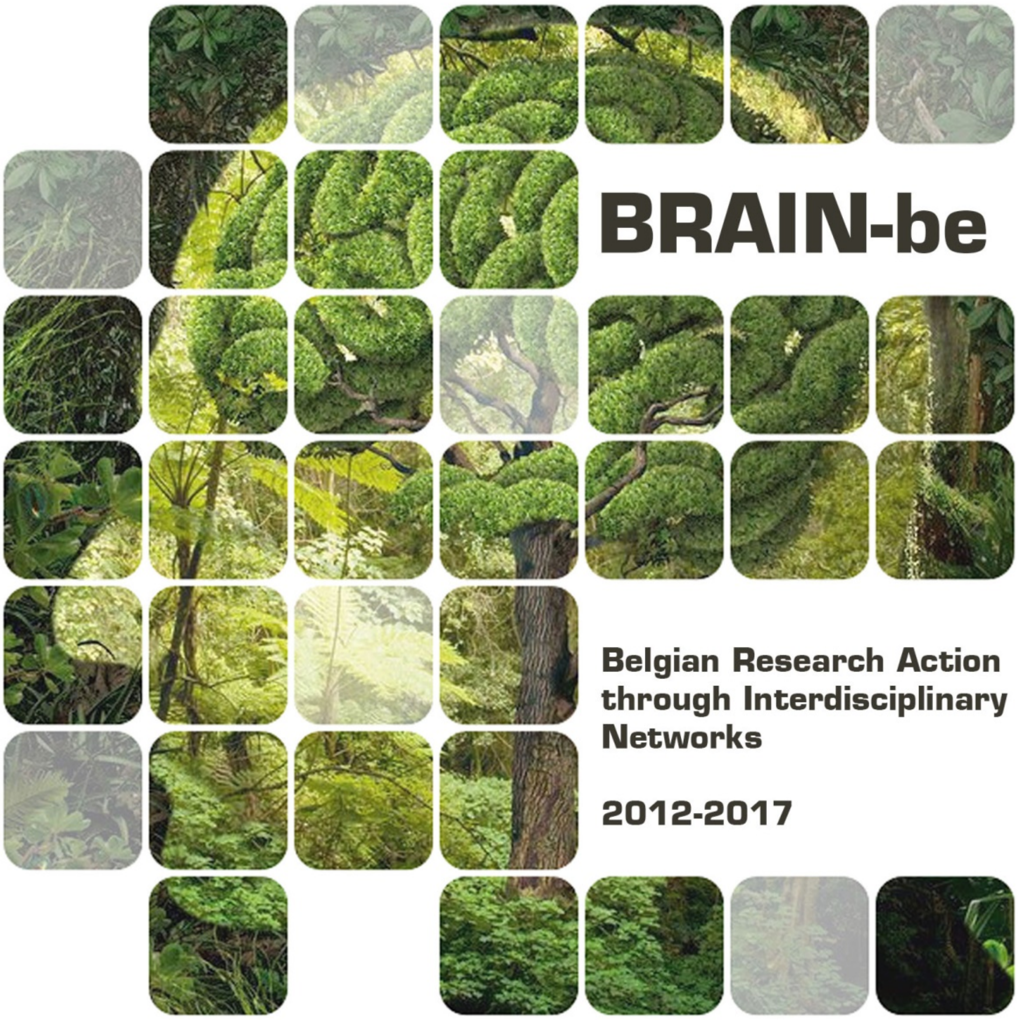

BRAIN-be

Belgian Research Action through Interdisciplinary Networks

2012-2017

VALUSUN Valorization of 70 years Solar Observations from the Royal Observatory of Belgium

LEFEVRE Laure (ROB) - DELOUILLE Véronique (ROB) - Von Sachs Rainer (UCLouvain) - MATHIEU Sophie (UCLouvain) - RITTER Christian (UCLouvain) - DUDOK DE WIT Thierry (Université d'Orléans) - CLETTE Frédéric (ROB)

belspo .be

NETWORK PROJECT

VALUSUN Valorization of 70 years Solar Observations from the Royal Observatory of Belgium Contract - BR/165/A3/VAL-U-SUN FINAL REPORT

PROMOTORS: LEFEVRE Laure (ROB)
VON SACHS Rainer (UCLouvain)

AUTHORS: LEFEVRE Laure (ROB) - DELOUILLE Véronique (ROB) - Von Sachs Rainer (UCLouvain) - MATHIEU Sophie (UCLouvain) - RITTER Christian (UCLouvain) - DUDOK DE WIT Thierry (Université d'Orléans) - CLETTE Frédéric (ROB)

Published in 2021 by the Belgian Science Policy Office
WTCIII
Simon Bolivarlaan 30 Boulevard Simon Bolivar
B-1000 Brussels
Belgium
Tel: +32 (0)2 238 34 11 - Fax: +32 (0)2 230 59 12
http://www.belspo.be
http://www.belspo.be/brain-be

Contact person: Maaike Vancauwenberghe
Tel: +32 (0)2 238 36 78

LEFEVRE Laure (ROB) - DELOUILLE Véronique (ROB) - Von Sachs Rainer (UCLouvain) - MATHIEU Sophie (UCLouvain) - RITTER Christian (UCLouvain) - DUDOK DE WIT Thierry (Univ. Orléans) - CLETTE Frédéric (ROB) - Valorization of 70 years Solar Observations from the Royal Observatory of Belgium - Final Report. Brussels : Belgian Science Policy Office 2021 – 43 p. (BRAIN-be - (Belgian Research Action through Interdisciplinary Networks))

TABLE OF CONTENTS

ABSTRACT

Context

The Sunspot Number is the longest scientific experiment still ongoing and a crucial benchmark to study solar activity, space weather and climate change. The Royal Observatory of Belgium (ROB) plays a central role in the continuation of this experiment, as it hosts the Sunspot Index and Long-term Solar Observations World Data Center (SILSO-WDC). This World Data Center aims at collecting solar data, as well as producing and distributing the International Sunspot Number, which is used in about two hundred scientific publications on an annual basis.

Objectives

The purpose of this project is to valorise two sunspot collections kept at ROB, which are presently neither available online nor exploited to the level of contemporary modern scientific standards.

The first collection consists of a series of about 20.000 digitized sunspot drawings acquired at the Uccle Solar Equatorial Table station (USET – images USET1940 and 2016) since 1940 and a corresponding database that is unverified, i.e. no complete quality checks could be performed, as of today. Moreover, the database is still incomplete in terms of extracted parameters. The digitization of these drawings started in the framework of a European project but lack of subsequent funding prevented us from adding essential metadata as well as achieving proper quality control. Thus, it was never made accessible to the wider community. As this dataset extends back to 1940, exploiting such a long-term and complete dataset is extremely important in order to assess the quality of parallel solar catalogues.

The second collection, the SILSO database, contains the numbers of spots and groups of spots on the Sun observed by a worldwide network since 1981 (more than 530.000 measurements – example image SunspotNumberDatabase.png). It is used on a monthly basis to compute the International Sunspot Number. Only one check is performed on each station every year, and its statistical basis dates back to the creation of the Wolf Number in the 1850's. A complete and consistent supervision of this database is important and as SILSO is the World Data Center for the determination of this index, it is our responsibility to bring the extracted Sunspot Number into the 21st century by exploiting this existing database to its full potential.

To this end, an international effort started in 2011 and focused on the past data, from the 19th century to the end of the 20th century. Unfortunately, the study made on this specific 1981-present dataset remained limited and modern standards can only be attained through the use of more modern statistical techniques.

The objective to achieve on our two collections is threefold: (1) Bring our databases to the modern era by adding essential metadata, whether it is additional parameters or techniques for quality assessment and quality control. (2) Use the value of solar parameters recorded in these databases to address today's scientific questions and (3) disseminate the collections and added value to a predefined set of audiences.

Conclusions

Collection 1: USET drawings

The USET-WDC-SILSO team and job students hired thanks to the project have been updating the database associated to the drawings throughout the project. This included data

quality control, metadata gathering and improving, the addition of a essential parameter to the existing list: the area of the groups recorded on the sunspot drawings. To this end, the development of a new software to analyse drawings was necessary as the previous one proved inefficient to determine sunspot areas on drawings.

Collection 2: SILSO database.

Thanks to the expertise of the UCLouvain partners, the SILSO database (1981-present) has been analysed and used extensively, in order to bring its exploitation to produce the International Sunspot Number into the 21st century. A full analysis of the statistical properties of the dataset has been realised and the results are published in Mathieu et al. (2019). Then a method to monitor (detect any anomalies in) the data has been developed and submitted to the Journal of Quality Technology (JQT: https://asq.org/quality-resources/pub/journal-of-quality-technology), and the PhD candidate Sophie Mathieu developed a full set of programs in Python so that the World Data Center SILSO can now apply the detection method in real time on the data (https://github.com/sophiano/SunSpot).

In parallel to this work, a full quality control of the input data in the database has been realised by the WDC-SILSO team, the corresponding metadata has been updated and the data has been made GDPR compliant (https://ec.europa.eu/info/law/law-topic/data-protection_en).

Keywords
Sunspot Number, Sunspots, statistics, error bars, monitoring

1. INTRODUCTION

Sunspots are dark spots appearing in groups on the solar surface. They are manifestations of solar magnetism and have been counted since the invention of the telescope in the early 17th century. In 1843, Samuel Heinrich Schwabe (Schwabe, 1844) discovered that time series of the sunspot count show a periodic pattern of approximately 11 years. This became known as the solar cycle, which deeply affects the entire solar system. The solar magnetic field embedded in sunspots is the driving force behind the solar variability that on a day-to-day basis influences the space environment of the Earth. In the mid 19th century, Dr. Wolf of the Zürich Observatory created a solar activity proxy by summing up the total number of sunspots with 10 times the total number of sunspot groups. With its 400 years, this “Sunspot Number” is the longest standing observational series of solar activity and has proven to be remarkably useful, not just for solar physics but in a large variety of physical sciences. Cross-correlation with other indices allows reconstructing the solar driving of Earth's Climate over the past few centuries. The sunspot number and the solar cycle are also a “Rosetta stone” for understanding solar-like variable stars. The digital library “The NASA Astrophysics Data System” collects more than 70 refereed papers per year containing the exact phrase “sunspot number” in the abstract or in the title.

Over the centuries, the Royal Observatory of Belgium (ROB) has played an essential role in the study of the Sunspot Number. Sunspot observations were regularly recorded since the early days of ROB (mid 19th century). Around 1908, Mgr. Eugène Spée (1843-1924), a Jesuit priest, started photographic observations of the photosphere at the ROB solar dome using a Grubb heliograph, which is still in use for the ROB visual sunspot drawings. In

1939 and during World War II, ROB started up a solar observing program in collaboration with the successors of Dr. Wolf at the Zürich Observatory. During 40 years, the counts derived from the ROB Sunspot Drawings were communicated to the World Data Center for the Sunspot Number at the Zürich Observatory. The original solar drawings, obtained at the ROB Uccle Solar Equatorial Table (USET) instrument, have been preserved and scanned and form a homogeneous series of uninterrupted drawings since March 1940. It is one of the longest stable observation collections of the Sun still active. In 1981, the Zürich Observatory ceased the World Data Center activity and in agreement with the international astronomical organization “COSPAR”, the World Data Center was transferred to ROB. As such, the ROB World Data Center “SILSO” (https://wwwbis.sidc.be/silso/) is responsible for collecting observations and counts of sunspots from an international and worldwide network of observing stations and for producing from these data the International Sunspot Number.

Today, USET is one of the last remaining operational telescope facilities at the Brussels/ Uccle site of the ROB and is a key-Belgian research infrastructure dedicated to solar physics. The collection of USET sunspot drawings and the SILSO collection of international observations, are unique datasets of scientific heritage. Despite their undisputed and long-standing value, these Belgian heritage collections of astronomical observations have only been used to extract in-house the International Sunspot Number by a procedure defined back in the Zürich era. The original observations have never been made available for external research, nor have they been explored with modern scientific methods and remain therefore below their maximum scientific potential.

Two collections are at the center of this project.

The first heritage collection, the ROB/USET sunspot drawings, consists of more than 23000 sunspot drawings acquired since March 1940. These drawings were originally made on paper but have been scanned and their scientific content has been largely digitized. This endeavor was paused due to funding limitations. In 2017, when VALUSUN started, the digitized drawings lacked contextualization and metadata and modern quality control was mostly missing. As a consequence, the USET drawing collection could not be made accessible to the scientific community nor used widely in scientific publications.

The second heritage collection is the database of the World Data Center SILSO that contains the numbers of sunspots and sunspot groups since 1981 provided by individual stations of the worldwide SILSO network and are used to compute the International Sunspot Number. Despite its success, the statistical stability of the International Sunspot Number needs constant concern. The ‘averaging’ process that combines the individual observations into one official International Sunspot Number was defined in the 1850s. The process depends on the choice of a reference to which an observation is compared. To this day, the reference is provided by a single station, called the ‘pilot station’, and as result the International Sunspot Number is sensitive to any problem that could arise with that station. Before the VALUSUN project, the performance of a particular station in reference with the pilot station was checked only once a year. A quality control of the observations of a station was done on an irregular basis. Therefore, the results of both these controls could not be shared with the observers or others. Even worse, the full content of the database itself was available neither to the scientific community, nor to interested people.

The central focus of this project was to open up the ROB Sunspot Data Collections to the international research communities and to contemporary scientific methods by contextualizing the collections with standardized documentation and metadata, updating them with new quality control mechanisms and making them accessible through an Open Access portal.

A multidisciplinary consortium has been set-up to achieve these goals, in which solar physicists at the Royal Observatory of Belgium collaborate with statistical experts at the Université de Louvain and with an international partner from the Université d'Orleans (France). The VALUSUN project was particularly cost-effective as only the young professionals in this consortium were charged to the project while the senior researchers at all 3 institutes worked in co-funding.

For the exploitation of our first collection, the USET drawings, we first extracted additional parameters through the development of a semi-automatic new version of the DIGISUN software. Extracting such parameters is paramount to the production of the real-time bulletins used in the alert system of the Regional Warning Center (RWC) of the SIDC. As a RWC, we need an alternative source of data to feed our alert system. Second, we assess the quality level of the existing and future associated metadata by comparing the extracted solar parameters catalogue to other overlapping catalogues. There are long term catalogues that are comparable but more heterogeneous, but thanks to the development of the multi-platform user-friendly DIGISUN 2.0 we can now compare USET data to the data from the pilot station of the SILSO network, i.e. a station that works with similar techniques and instruments and the same extraction software, making it the best candidate to evaluate the quality of the USET data.

For the exploitation of our second collection, the SILSO database and the sunspot counts it contains, we use advanced techniques from applied statistics. It is at the foundation of data analysis, a discipline which uses different kinds of methods to extract information from data. Hence, its two fundamental ingredients are the data and the techniques that can be used for cleaning, modelling or processing those data in order to find new evidences or support decision-making. We have two main goals for this collection.

The first one is analysing, modelling the errors and monitoring the quality of the sunspot counts over time. To this end, we develop a comprehensive error model for the sunspot numbers in a multiplicative framework. The model decomposes the data into a physical signal, common to different observers, corrupted by three types of errors, at short-term, long-term time periods as well as during solar minima. We provide a complete analysis of the different terms of the model, including parametric fits of their distributions. This model allows us to obtain more robust estimators of the sunspot numbers and to provide errors for those data at each point in time. It also highlights the long-term deviations that occurred in the past series of several observing stations. Although specially adapted to the sunspot numbers, the model may also serve as a source of inspiration for treating other datasets with similar properties.

The second main objective of this work is the construction of general non-parametric methods to monitor the sunspot counts. We construct a non-parametric monitoring procedure based on control chart and support vector machine to efficiently detect the deviations of sunspot numbers over time. The scheme is designed to work with non-normally distributed and autocorrelated processes with potential missing values. It works at different scales and does not require any parametric assumption about the data to correctly operate. This method allows us to automatically identify many deviations in the sunspot numbers, mostly unseen in previous analyses and helps us to find the root-causes of some prominent shifts. This control scheme will be implemented to monitor all observing stations involved in the counting process, to prevent the future build-up of large deviations over time, such as those previously observed. As this is a growing database that still receives data on a daily basis, the process of determining a multi-station reference needs to be done dynamically. The current processing, done every month and every 3 months, is a historical heritage based on a very slow availability of data that is no longer relevant. Hence, those novel statistical methods need to be rendered dynamic in order to provide a method that is computable in near real-time.

This project is at the crossroads between Solar Physics, scientific programming and mathematical statistics. In addition to that, its output presents a huge interest for domains outside of Solar Physics such as Space weather and Climate Science. Because of the importance to these disciplines, the modern evaluation of the Sunspot Number and the use of our local data for warning bulletins could have a non-negligible impact on decision-making.

We ensure a wide dissemination of our final products. The drawings and their metadata are made available to the general public and a last quality check is performed on these data in the context of a “citizen science” project. The sunspot number database is also available: in particular, each observer of the SILSO network will have access to his or her real-time quality assessment delivering a much needed feedback to our faithful observers.

2. STATE OF THE ART AND OBJECTIVES

State of the art

The techniques behind sunspot drawings and counting have been static for many decades. Despite the advance of CCD-equipped telescopes, sunspot drawings have been deliberately kept unchanged to continue historical time-series in similar circumstances as centuries ago. In this sense, the ROB Sunspot Data Collections truly are a historical heritage.

Because of its importance in the context of space climate, space weather (solar storms) and even the Earth climate (evolution of the total solar irradiance) the oldest solar index, also called the International Sunspot Number, has been the center of attention of scientists all over the world, but especially here, at the World Data Center SILSO, its curator for the last 50 years. Increasing concerns on the stability of this important index have sparked a strong increase in the research activity on the process from raw sunspot observations to an internationally agreed, stable solar activity index. A series of international workshops (Cliver et al. 2015) have been organized to correct historical discrepancies in the calibration of the International Sunspot Number. Such discrepancies could be identified by studying changes in the ratio of the sunspot number with similar indices, such as the Group Sunspot Number (Hoyt et al. 1994, Hoyt & Schatten 1998a,b). By relating these discrepancies to known transitions in the circumstances in which the sunspot number was produced, it has been possible to recalibrate specific incoherencies in the sunspot number (Clette et al 2015, 2016).

To avoid a multiplicative accumulation of historical errors, a “Backbone” method (Svalgaard 2013, Chatzistergos, 2017) was constructed based on overlapping, “high quality” sunspot observers with particular long time spans. Going even further, a method that completely avoids the need for overlap was developed in 2016 (Usoskin et al., 2016) and refined over the last few years (Willamo, 2017, 2018, Usoskin, 2021). Obviously, in this process it becomes clearer that proper contextualization and documentation of all the available data are crucial.

The international effort to recalibrate the Sunspot Number is not finished yet. In the last few years, new questions came up. Observations point out that the magnetic field strength in sunspots has had a secular decrease from 1998-2012 (Livingston et al. 2012, and references therein). In addition, after the solar minimum of 1996, the strong correlation between the International Sunspot Number and the solar radio flux at 10.7 cm (e.g. Kundu 1965) appeared to break down (Svalgaard & Hudson 2010). Also, the long-term ratio of the number of spots to groups has decreased from its nominal value of 10 (Clette et al. 2014). These issues also lead to new questions on the confidence level of the International Sunspot Number and the “quality” of each contributing station in the SILSO network. The SILSO WDC considers it as a must that all the research it produces should be transparent and reproducible by independent researchers. The whole science community should be able to participate in this upcoming research, not only the SILSO World Data Center. In parallel to the VALUSUN project, a joint effort to develop and validate techniques to stitch data together has been going on (ISSI), and also stresses the need to make all data available and contextualised (i.e. metadata). The contextualisation of historical sunspot data from all over the world is at the center of the FARSun project submitted to BELSPO in 2020.

As stated previously, the ISN still suffers from a number of historical errors and inconsistencies. Some of them have been partly addressed by the recalibration of the ISN in 2015 (Clette and Lefèvre, 2016). Even the most recent part (1981-now) of the series lacks however a proper error modelling. The WDC-SILSO team is currently working on improving the ISN computation and coordinates an important community effort to correct past errors (ISSI, sunspot workshops). This project addresses the specific problems of the most recent part of the series, contained in the WDC-SILSO database.

From the statistical point of view, the main contributions of this work are therefore two-fold (one-fold for the data and one-fold for the method). Our first contribution is related to the specific dataset that is studied here: the sunspot numbers. Our main objective is to go beyond the aforementioned historical heritage and propose a thorough statistical treatment of the building blocks of the ISN: Ns, Ng and Nc. This include the development of a comprehensive uncertainty model for the sunspot numbers as well as the construction of a robust monitoring procedure for supervising the quality of these numbers over time. Then, the second main contribution of this work is related to the development of a general non-parametric monitoring procedure, which can be applied to different panels of time-series for quality control.

As an index derived from count data, Nc (or Ns and Ng) does not necessarily follow a Gaussian distribution (Vigouroux & Delache 1994; Usoskin et al. 2003; Dudok de Wit et al. 2016). A processing based on sigma-clipping as it is done now is thus not fully adapted, but still undoubtedly better than what was done during the Zurich era. Long-term analyses started with models of the shape of the sunspot number time series (Stewart & Panofsky 1938; Stewart & Eggleston 1940). They pursued the works by M. Waldmeier himself (Waldmeier 1939), who tried to understand the solar cycle and predict upcoming cycles. Later on, Morfill et al. (1991) investigated the short-term dynamical properties of the SN series using a Poisson noise distribution superimposed on a mean cycle variation. Vigouroux & Delache (1994) also use a Poisson distribution to approximate the dispersion of daily values of the SN at different regimes of solar activity. Usoskin et al. (2003) develop a reconstruction method for sparse daily values of the SN and model the monthly number of groups corresponding to a certain level of daily values by a Poisson distribution. Schaefer (1997) emphasizes the need for error bars on the AAVSO sunspot series (Foster 1999), and more recent results in Dudok de Wit et al. (2016) present a first uncertainty analysis of the short-term error, through time domain errors and dispersion errors among observing stations, still assuming a Poisson distribution. In Dudok de Wit et al. (2016), however, the authors uncover the presence of overdispersion in the SN and approximate the SN by a mix of a Poisson and a Gaussian distribution in an additive framework. Although non-Poissonian, this additive model fails to capture some of the characteristics of sunspot data. Chang & Oh (2012), on the other hand, use a multiplicative model to simulate sunspot counts in view of assessing the dependency of correction factors on the solar cycle.

Objectives

The overarching goal of the VALUSUN project is to make the ROB Sunspot Data Collections accessible for worldwide external researchers. In particular we want to make the USET sunspot drawings available as a “high quality” observation station from 1940 till present, with proper contextualization and documentation to make it a primary reference (ref. Backbone method). The goal also extends to opening up the SILSO database of individual sunspot count observations to investigations of external researchers and provide state-of-the-art quality monitoring of the stations.

To achieve this goal, we use the following objectives as guiding principles:

- The shared ROB Sunspot Data Collections must be as complete as possible, i.e. they must retain a maximum of the scientific content that was available in the original data.
- The quality of the long-term data sets must be monitored continuously and the results of the quality checks must be accessible to both the data providers and the user. In case of quality loss, the affected observation station must be alerted.
- The ROB Sunspot Data Collections, together with the metadata, full context and documentation must be readily (online) available following an Open Access philosophy.
- The science communities, for which the ROB Sunspot Data Collections are relevant, must be made aware of the updated ROB Sunspot Data Collections.

3. METHODOLOGY

A number of methods were needed to achieve the goals and objectives described above for both ROB Sunspot Data Collections. We first describe the initial workplan designed to achieve our goals within the designated timeframe and then the different scientific methods necessary to achieve the different goals.

Initial workplan and Detailed description of the tasks and subsequent adaptations

The work breakdown structure of this project is aligned with the main aspects of the call, namely “updating, contextualization and accessibility for research of [...] a set of collections”.

There are 3 main work packages in addition to the management of the project (WP1). WP2 is “Contextualization of the ROB Sunspot Data Collections”, WP3 “Updating of the ROB Sunspot Data Collections” and WP4 “Accessibility of the ROB Sunspot Data Collections”.

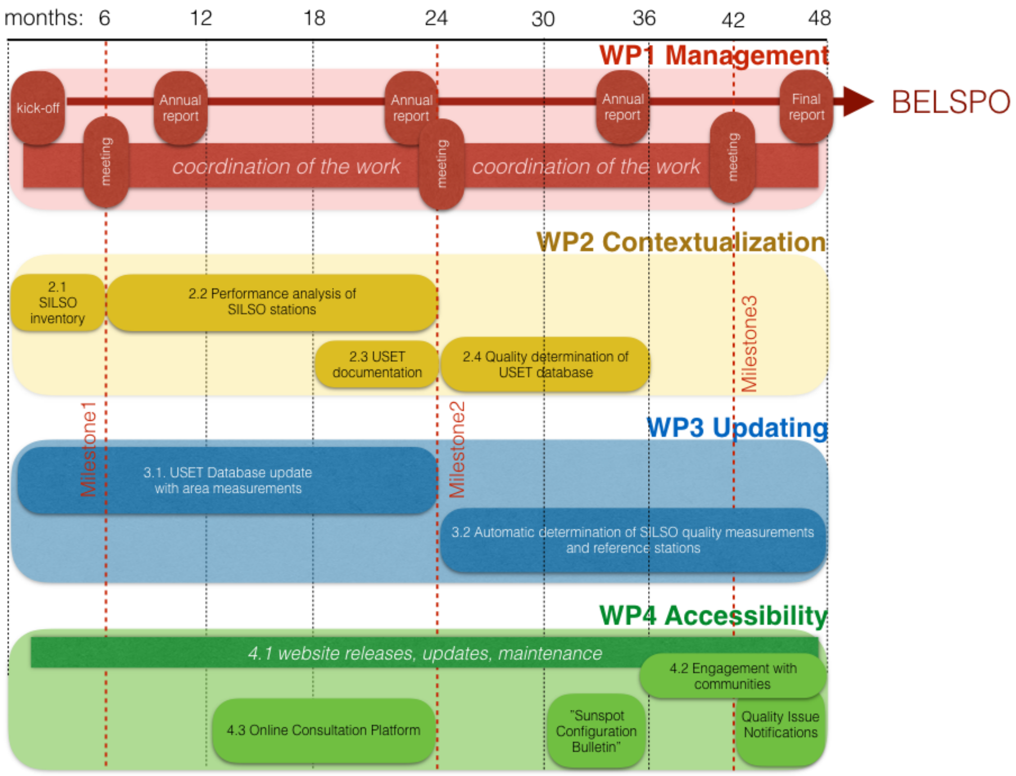

Fig. 1 Initial work breakdown structure and overall planning of the VAL-U-SUN project

WP1: Management (WP manager: ROB/Lefèvre)

The project coordinator Dr. Laure Lefèvre is responsible for the overall management of the project and take up all the roles foreseen in the BRAIN-BE-Call2016 document, i.e. interfacing with BELSPO (1.1) and coordination of the work (1.2).

WP2: Contextualization of the ROB Sunspot Data Collections (WP manager: UCLouvain/von Sachs)

For datasets covering decades, it is essential to record and monitor the circumstances in which the observations were made and to follow up the evolving quality of the observations. Understanding and analysis is greatly facilitated if this context is readily available and accessible to the researcher. In this work package, we work on the contextualization of the ROB Sunspot Data Collections by standardizing and extending the available documentation and metadata by carefully recording the quality of the time series and discrepancies therein.

Table I. Planning of the tasks of WP2

Task		planned	realized
2.1	SILSO network documentation	M1-6	M1-12
2.2	performance analysis of SILSO stations	M6-24	M6-36
2.3	USET stations documentation	M18-24	M18-24
2.4	USET Quality control	M24-36	M24-48

Deliverables: Open Access Website will provide the available documentation for WDC-SILSO stations (2.1), performance analysis of said stations (peer reviewed article, 2.2), USET station documentation (2.3) and the analysis of the quality of the USET data (peer reviewed article, 2.4).

Adaptations: As can be seen from Table II, as the tasks could evolve in parallel thanks to our colleagues in UCLouvain, it was possible to adapt the planning to extend until the end of the project.

WP3: Updating the ROB Sunspot Data Collections (WP manager: ROB/Clette)

The ROB Sunspot Observations Collections are enriched with new information derived from state-of-the-art statistical and software methods, as well as from the results of the previous WP (contextualization). This update opens up these heritage data collection for modern exploitation through Open Access by the broad science community.

Table II. Planning of the tasks of WP3

Task		planned	realized
3.1	USET area measurements	M1-24	M1-48
3.2	quality and reference station SILSO	M25-48	M25-48

Deliverables: USET (3.1) and WDC-SILSO (3.2) databases made available for external research (M48).

Adaptations: In this work package, the USET area measurement task, realized by job students was extended until the last year of the project (June 2020) because of a lack of students the first year and the excessive unfriendliness of the area measurement module of DIGISUN 1.0 that led to the development of DIGISUN 2.0 in 2018.

WP4: Accessibility of the ROB Sunspot Data Collections (WP manager: ROB/Lemaitre)

The ultimate goal of this WP is to make the ROB Sunspot Data Collections openly accessible for modern research. Project results are published online (Task 4.1). One-way accessibility in itself is however not sufficient, we aim at a more interactive engagement with our target user communities (Task 4.2). Finally, there is a specific task for the creation of an online consultation platform and the analysis of the resulting feedback (Task 4.3).

Table III. Planning of the tasks of WP4

Task		planned	realized
4.1	Open Access data and documentation website	M1-48	M1-48
4.2	Engagement with Scientific Communities	M25-48	M25-48
4.3	Online consultation platform	M13-24	M13-48

Deliverables: The WDC-SILSO website has been extended with project deliverables (4.1), the results of VALUSUN were presented at relevant scientific meetings (4.2), and an online consultation platform has been set up (4.3).

Adaptations: The biggest change concerns the Online consultation platform. Although it was set up on schedule, the number of visits remained limited (109 persons participated in the proposed exercise). Consequently, the results are difficult to interpret on a larger basis, but they are presented in this report.

Whereas the Work Plan gives a complete overview of all the tasks and methods employed, we focus now on the details of the more technical methods.

Methods for Sunspot drawing digitization

At the beginning of the VALUSUN project, the USET station used a software application for sunspot drawing digitization called DIGISUN (we will call it DIGISUN1.0 from now on). The application allowed an operator to extract from a drawing the number of groups on the Sun, their positions, the number of spots inside each group, their Zürich and McIntosh classifications, and their dipole tilt. A first processing of the USET sunspot drawings with DIGISUN1.0 was done during the FP7-SOTERIA project between Nov 2008 and Oct 2011 but was never completed: at the beginning of VALUSUN, 10% of the drawings had not been digitized.

Meanwhile, an improved DIGISUN1.0 application had been prototyped in 2016 as an added area-measurement module. Using a semi-automatic technique, including masking and region growing within an operator-defined region, pixels are counted that correspond to the sunspot area. With the location of each pixel known, the spherical projection effect is corrected to provide a more consistent time series of the area during the solar rotation.

Originally, we planned on exploiting this area module within VALUSUN but quickly realised it was not efficient enough at computing areas on a large scale, i.e. not user friendly.

So between 2017 and the beginning of 2018 a whole new DIGISUN2.0 software was developed, fully implemented in Python and compatible with all platforms and OSs. As we are talking about a lack of efficiency in DIGISUN1.0 compared to DIGISUN2.0, all the existing database was not recomputed. Only day-to-day drawings and areas of older drawings (one task of this project) were computed with DIGISUN2.0.

The resulting database of sunspot group characterization underwent a critical quality assessment using techniques developed for the analysis in Lefèvre et al. (2011, 2014), as well as techniques developed with colleague Maarten Janssen from ULB, and PhD student Shreya Bhattacharya (ROB-ULB). The report is presented in the *scientific results* section below.

Methods for quality measurement and creation of reference stations for the WDC-SILSO network

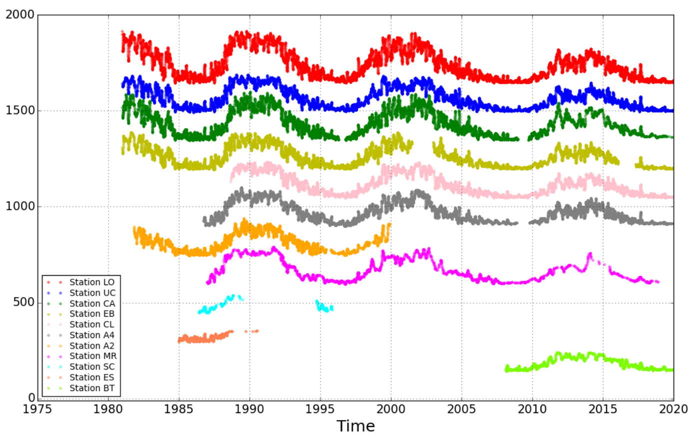

Figure 2. Sunspot number (vertical axis) as derived from the data available for single stations, as a function of time (horizontal axis).

The WDC-SILSO database contains sunspot counts from 300 observing stations (as of October 2021) that contributed between 1981 and the present day, but at any given time, only a subset of stations contribute data. Fig. 2 presents the raw sunspot numbers (monthly averaged) given by 11 different stations. The (upper) red curve shows the data of the Locarno station, which is used nowadays as the pilot or reference station. The (second) blue curve corresponds to the sunspot numbers from the USET station in Uccle. This figure highlights one of the main challenges we are facing, which is the high inhomogeneity of the database. A significant effort is dedicated in this project to research the quality of the contributions of each station and the implementation of quality measures. By quality, we mean long-term and overall stability of the observations within the network.

Most sunspot records lack confidence estimates, which are essential for performing statistical tests and for doing quality assessment. To make up this shortfall, we recently developed a series of methods for estimating the precision, out of which confidence estimates can be derived (Dudok de Wit et al. 2016). For example, one method consists in fitting a model of an inhomogeneous Poisson point process to the data per station and subsequently calculating a standard deviation for each station at every day. These methods will be adapted to the databases and applied in a systematic way to all records.

In VALUSUN, we propose the first systematic and thorough statistical approach for monitoring the sunspot numbers. The method consists of three steps: smoothing on multiple time-scales, monitoring using block bootstrap calibrated CUSUM charts and classifying of out-of-control situations by support vector techniques. Many different procedures have been developed in the Statistical Process Control literature to monitor the mean of univariate processes (Qiu, 2013; Montgomery, 2004). Some of them such as the classical cumulative sum (CUSUM) chart (Page, 1961) are promising but cannot be used directly here because the mean and variance of stations changes over the solar cycle Indeed, if each station is an univariate process, its mean and variance change over time (Hathaway, 2010). To compensate for this, we use a method based on the dynamic screening system developed by Qiu and Xiang (2014). To the best of our knowledge, this method is the only one that can be adapted to the particular characteristics of the sunspot data: the non-normality, autocorrelation and non-stationarity of the data as well as the absence of "in control" (typical) periods in all series.

Furthermore, since many aspects of the method are general, the monitoring scheme can also be applied to a much wider range of problems.

METHODS concerning ACCESSIBILITY

The ultimate goal is to make the ROB Sunspot Data Collections fully accessible for modern scientific exploitation. All the original data, as well as all results of the project will be accessible online following the ROB open access data policy (see http). One-way accessibility in itself is however not sufficient, we set up a more interactive engagement with our target user communities, hence registration will be asked before accessing the data distribution section of the website. In particular we foresee to actively communicate the results of the project through :

- the International Astronomical Union’s (IAU) Inter-division B and E working group on Coordination of Synoptic Observations of the Sun (through the follow-up Committee)
- the series of International Workshops on the Sunspot Number (http://ssnworkshop.wikia.com/wiki/Home, Cliver et al 2015)
- the annual European Space Weather Week (see http://esww17.iopconfs.org/Home for this year’s edition)
- the ROB communication channels (e.g. http://www.stce.be, http://www.astro.oma.be/en/)
- As deliverables of the project, we will also issue direct email-messages to specific target groups

Finally, we built an online consultation platform. This platform (essentially a part of the VALUSUN website) shows a number of sunspot drawings and the participants are asked to count the number of spots and groups. The idea is to assess how people with different levels of expertise interpret a sunspot drawing in terms of the sunspot number. As the sunspot number is the sum of the number of sunspots and 10 times the number of sunspot groups, the identification of sunspot groups is critical. We invite participants from three groups with different expertise level: expert solar physicists, SILSO network observers, and the general public. We will study how the interpretation varies from one participant group to the other, e.g. is there a difference between the participants groups and if so, are the SILSO observers (many of whom are amateurs without training) interpreting drawings like expert solar physicists or more like the general public? This investigation is motivated by the realisation that early sunspot observers did not have any prior expertise and thus probably resembled the general public of today. Also, as a SILSO observer gains more expertise, (s)he might shift from one category to the other and thereby increase the quality of his/her interpretations.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

Expected research results

The research within this project is focused on new methodologies and techniques for a better quality control, and a better accessibility to the information buried in the ROB sunspot data collections.

Among the highlights, we expected the following research results:

- Open Access publication of more than 70-year long coherent collection of sunspot observations from the USET facility, one of the last active telescope platforms at the ROB site in Uccle. The data set will be fully digitized, documented and contextualized.
- Besides the sunspot number that can be derived from the above dataset, we will extract a new characterization of solar activity from the USET drawings, namely the sunspot area. Also this new parameter will be fully validated and available for external researchers.
- Open Access publication of the SILSO database with a collection of more than 35 years of sunspot number observations contributed by 285 observing stations worldwide. The database will be fully documented and contextualized, including modern quality control techniques on individual stations.
- An automated procedure will be set-up to identify a set of high-quality observing stations that together can form a new multi-pilot reference group.

We also expected the base research activities supporting the above developments to be published in peer reviewed journals, which has been done for the SILSO dataset (Mathieu et al. , 2019, Mathieu et al., 2021) and in progress for the USET dataset (Aparicio et al., 2018). The statistical work on quality control has been embedded in a PhD thesis that will be held on Nov. 9th 2021. Finally the work on specific e-mail messages (“Quality Issue Notifications”, “Sunspot configuration Bulletin”) to disseminate research results in near-real time is still a work in progress. All the base material for these bulletins exists, i.e. the data to be included have already been computed and tools developed to extend them to all stations, but the programs developed by S. Mathieu have to be encapsulated within the WDC-SILSO framework for the actual publication of these bulletins.

Achieved Work (Results)

WP2:

Task 2 .1:SILSO network documentation and metadata inventory

With colleague O. Lemaître, we started an inventory of which information was already collected from our network of observers. We launched a campaign to check the email addresses of all our current observers. Previously, information on the instrument and

technicalities were stored in a single column as comments. They are now separated as different kinds of information in the database.
We now have in the database:

- Station Prefix: two letters
- Station Number: 3 digits
- Login/Pasword/Access Date (access date: date of creation in the DB)
- Precise name of Institute or Observatory/Name/Description
- Person of contact/Email
- Type of instrument (reflector/refractor)
- Diameter, Focal length
- Type of mount/Filters
- Projection/Eyepiece (size of drawing/eyepiece magnification)
- Number of observers
- Precise coordinates (Latitude- Longitude)
- Postal address
- Second/Backup Email
- Country

The inventory of which information was already collected from our network of observers is now done. We checked the email addresses of all our current observers. In addition to the work planned for VALUSUN, we also endeavoured to make the SILSO database and especially the metadata, GDPR compliant, which is why this task took longer than expected.

Present status: Some information concerning older observers (no longer observing) from remote locations could not be retrieved at this time because contact information is not valid anymore. We dedicated the time allocated by the VALUSUN project to look for them through different means, i.e. via the internet, but could only retrieve a small part of the information, especially for remote locations with different languages.

Recommendation:

- *F*or future projects, we would recommend having experts collaborating who speak a large set of languages to facilitate the search for information in remote locations for french or english speakers.
- Also, it stresses the need for long-term investment in the maintenance of the accuracy of the databases. In this context, the presence of an additional operator to maintain contact with older sunspot number providers and/or the development of an interface (social networks, blog, forums etc…) to maintain or initiate contact between past, present and future observers as well as with the WDC-SILSO. The WDC-SILSO curates and produces data that generates about 100 publications per year, and does not have its own funding from the government, an under-funding that sometimes leads to loss of data.

Task 2.2: Performance analysis of SILSO network stations

To study the stability of the Sunspot Numbers of the different observing stations of the SILSO network, we first needed to develop a model for the Sunspot Number. In that context, we organized different meetings between the different partners and we met with members of the follow-up committee during the course of year one.

First, the recruiting of a PhD taking some incompressible time (PhD students usually start their work with the academic year in September), we devised an alternative plan to progress at the beginning of year one and obtain results we could build on with the future PhD student.

Based on the article by Dudok de Wit, Lefèvre & Clette (2016) analysing the errors in the Sunspot Numbers, students under C. Ritter from the Statistical Methodology and Computing Service (SMCS) analysed different aspects of the signal during the period Mar-April 2017. The students worked on an approximately 40-hours project.

There were two weeks of exploration of the subject, after which the students were organized in 3 groups that worked on different aspects of the problem. This includes (1) how to fill in gaps in the data, (2) how to study the errors in the sunspot numbers (stabilize the variance of the data, model the evolution of the sunspot number from one day to the next, dispersion error) (3) how to quality control the different series. They reviewed the variability and errors of the different Sunspot Series that enter into the International Sunspot Number (SILSO). They evaluated Haar-Fisz transform versus Anscombe transform (or more simple square-root transform) of the series to stabilize the variance. They evaluated a series of ARIMA models to better represent the Sunspot Number errors (one-day-ahead errors). To realize the quality control of the stations they used CUSUM methods, or EWMA (Exponential Weighted Moving Average). They confirmed Dudok de Wit et al. (2016) results that observational errors (dispersion errors) are mostly Gaussian while time domain errors (determined with ARIMA models) are mostly Poissonnian.

They also translated different programs given by T. Dudok de Wit in the context of our VAL-U-SUN collaboration, into other languages giving the future PhD student a strong programming base. Presentation of the work was done at the end of April 2017, and reports produced by each student thereafter. These reports are available in the annexes of the present report. The student project was organized at no cost for the project, but represents a great outcome, and was a building block for our PhD student.

Task 2.3:USET documentation

The inventory of historical information available at ROB is complete.

Task 2.4:USET quality control

The largest part of the USET quality control has been realized during Year 4, thanks to the development of DIGISUN 2.0. However, part of it was done as soon as 2018 in the article published with Aparicio et al. (2018). The more scientific results of the quality control of the USET database are presented in the next section.

WP3:
Task 3.1: Update of the USET database with area measurements (due month 24 but started month 1)

13 months of job students were hired between June 2017 and September 2020. The initial number was estimated at 12, but there was a remnant in the budget in 2020 that was transferred to hire 1 more student. Overall, the initial estimation of the work to perform was good. We hired 3 students in 2017, 5 in 2018, 3 in 2019 and 2 in 2020.

During the summer 2017, job students worked on the DIGISUN interface to add area measurements to drawings already digitized. Before the start of the VAL-U-SUN project, the areas where only computed on a day-to-day basis by the USET operators whenever they made a daily drawing (started in July 2017). With VALUSUN, we added personnel working specifically on adding area measurements to all digitized drawings.
During 2018, we hired 5 student/months for summer jobs to make up for the lag in area measurements due to less hiring of job students in year 1.
To improve our ability to fill in the database, our colleague S. Bechet developed a whole new version of the Digisun software to make up for recurrent and time-consuming problems in the old version.
During 2019, we completed all possible years considering the manpower assigned through job students. We realized a manual quality control on 3 solar cycles at the beginning of year 3, and planned a full-sized quality control on the database for 2020, and a complete quality control after the summer.

During year 4 (2020), 2 job students finished the remaining parts of the digitization.

WP4:

Task 4.1: Website.

The VAL-U-SUN website is online. Results are published jointly on the SILSO website and VAL-U-SUN website. The open-access VAL-U-SUN website presents the results of this project whenever they are available. A first version of the network access to their metadata and the quality analysis exists at http://www.sidc.be/valusun/wolf/ We give access to all observers to their own metadata (not the metadata of the other stations for privacy reasons), and our analysis of their quality. The first version of this quality control process is the computation of a monthly k-factor that they can access from the website.

Task 4.2: Engagement with scientific communities.

The VAL-U-SUN results where published and presented at each meeting.

Task 4.3: Online Consultation Platform.

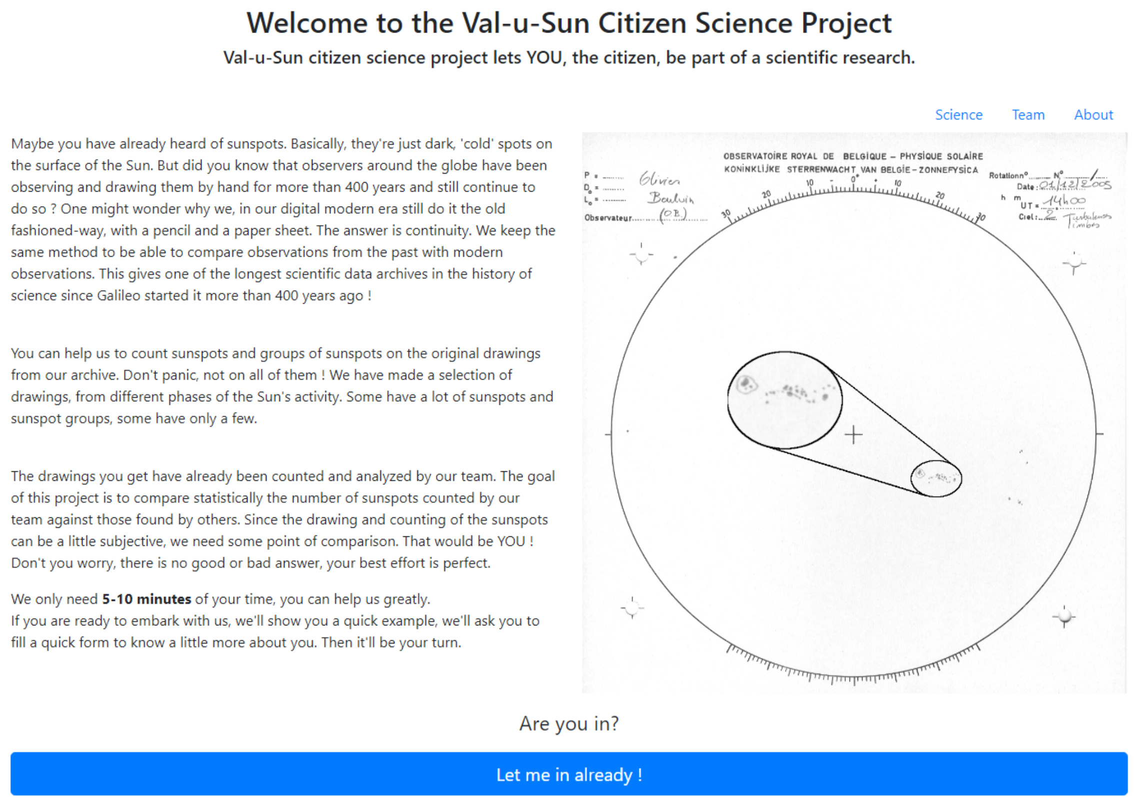

Figure 3 Screenshot of the VAL-U-SUN online consultation platform developed by O. Lemaître

The VAL-U-SUN online consultation platform is online (Figure 3). Its development started in 2018, but it came online at the beginning of 2019 and results started coming in late 2019.

There were 109 contributors to the project. We took all contributors that analysed 3 or more drawings and looked at the difference between their estimations of the numbers of spots and groups and the USET estimation from the database.

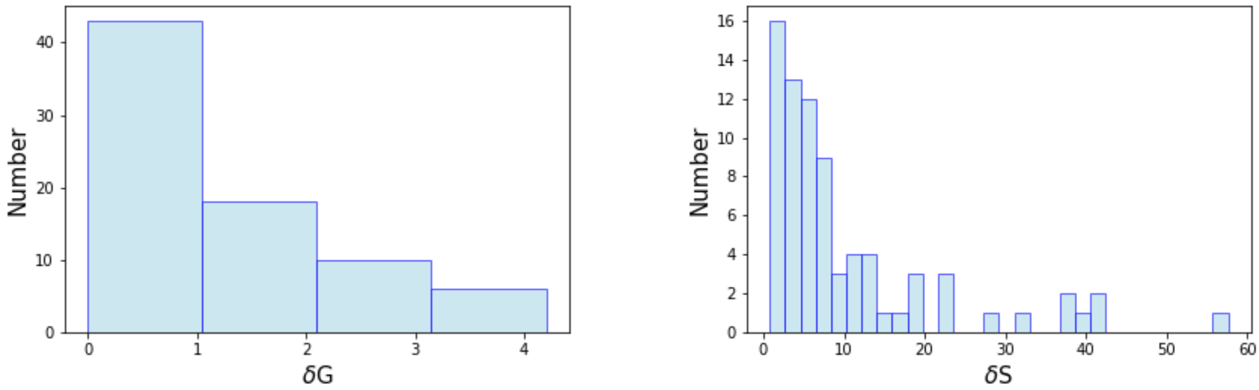

Figure 4. Results of the VAL-U-SUN online consultation platform : distribution of the difference between the USET database and the estimation of the number of groups (deltaG) and spots (deltaS) for all contributors having worked on 3 drawings at least.

Although the total number of observations is low (109 contributors), we can see in fig. 4 that, even without any background, most people identify the same number of groups. This is probably due to the fact that most drawings are not very complex cases. The identification of the number of spots seems to tell a slightly different story. Considering the statistics however, it is impossible to conclude at this point, but it is interesting to note, that, contrary to current beliefs, counting the number of groups seems less prone to errors than counting the number of spots.

Remark: Although the number of contributors was too small for a large-scale study (109 contributors), the development of this platform enabled ROB to acquire the skills to develop such citizen science initiatives and we now know that additional channels of communication are needed as well as more frequent publicity to a wider public.

Recommendations:
In this consultation, we invited people to log in relevant information like their age and eyesight for example. Although this was for scientific purposes, it is possible that the amount of information asked was a negative factor for part of the population. This is something that we will definitely keep in mind in the future when we “consult” citizens on scientific problems. In addition to that, we will keep on improving and exploiting the results of this consultation platform, as we will use whatever manpower we have at our disposal to maintain it for as long as possible.

Scientific Results

WP2

Task 2.2: Performance analysis of SILSO stations

Development of a Model of the number of spots (Ns) and number of groups (Ng), as well as the composite (Nc=10 x Ng +Ns) , for the analysis of the Sunspot Number

As an intermediary result to this specific work package of the VAL-U-SUN project, I mention here the reports by the students from IASB/UCLouvain in the context of their consulting course, available upon request to the coordinator.

Another product from our project is the article by Sophie Mathieu, Veronique Delouille and the VAL-U-SUN team that has been submitted for the 2018 IEEE Data Science Workshop. The 2018 IEEE Data Science Workshop is a workshop that aims to bring together researchers in academia and industry to share the most recent and exciting advances in data science theory and applications.

In the context of our project an **article by Sophie Mathieu, and the VAL-U-SUN team** was published in the Astrophysical Journal (Mathieu et al., 2019).

In this article, we build upon the work in Dudok de Wit et al. (2016), which presents a first uncertainty analysis of time domain errors and dispersion amongst the stations assuming a Poisson distribution. In the present paper, we propose a more comprehensive error model that accounts for all types of errors known to the experts, and fit them with distributions that comply with essential features of the number of spots (Ns) and the number of groups (Ng), namely their zero-inflated and overdispersed nature. We present the dataset, a proxy for the *true* solar signal, which is further used to estimate the various errors and our results on error estimation.

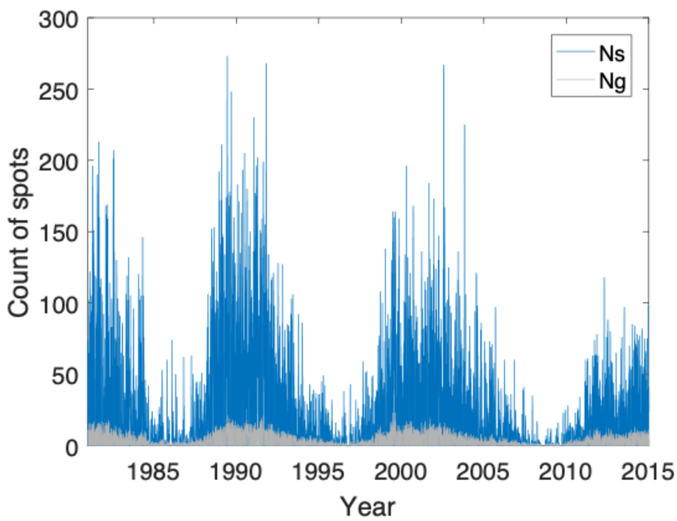

Figure 5: Daily number of spots (Ns) and number of groups (Ng) observed in Uccle (USET facility).

Having such a systematic quantification of uncertainties is crucial for properly building composites out of multiple observations and for monitoring the stability of stations over time.

We propose the **noise model:** $Y_i(t)=(\varepsilon_1(t)+\varepsilon_2(i,t))\,s(t)+\varepsilon_3(t)$ **(1)**

Where $Y_i(t)$ represents either the number of spot counts Ns or the number of group counts Ng recorded by station i at time t. The term s(t) is a random variable (r.v.) representing the *true* solar signal, i.e. the actual number of spots or sunspot groups. We denote E (s(t))= $\mu_s(t)$ where E is the expectation sign.

Results in Dudok de Wit et al. (2016) evidence a **short-term**, rapidly evolving, dispersion error across the stations that accounts for counting errors and variable seeing conditions: we denote it **$\varepsilon_1(t)$.** We introduce **$\varepsilon_2(i,t)$** to handle station-specific **long-term** errors such as systematic bias in the sunspot counting process that we call a **drifts** estimated thanks to **$\mu_2(t)$**.

Both ε_1 (t) and ε_2 (i,t) are multiplicative errors, as typically a station will count x% more (or less) sunspots than a reference station. Finally, an additive error term, identically distributed (i.d.) amongst the stations, is denoted $\varepsilon_3(t)$. It models essentially errors occurring during minima of solar activity, when there can be extended periods with no or few sunspots. ε_3 thus captures effects like short-duration sunspots and non-simultaneity of observations between stations.

Figure 5 displays raw Ns and Ng observed at the USET solar station of the ROB. The short-term variations coming from the solar variability and the observational errors are clearly visible, superimposed on the imperfect seasonality of the eleven-year solar cycle and the possible drift of the station.

To differentiate ε_1 from ε_2, we use the fact that they live on different time scales: ***ε_1 will be estimated on short time scale (< 27 days)***, while the ***mean of ε_2 will be estimated when a low-pass filter is first applied on the time series*** (short time scales are filtered out).

For time scales smaller than 27 days, ε1 is dominant in Equation (1), whereas ε2 should not contain any drift (on short timescales no drifts exist). With $\varepsilon(t)=\varepsilon_1(t)+\varepsilon_2$, Eq. (1) reduces $Y_i(t)=\varepsilon(t)s(t)+\varepsilon_3$

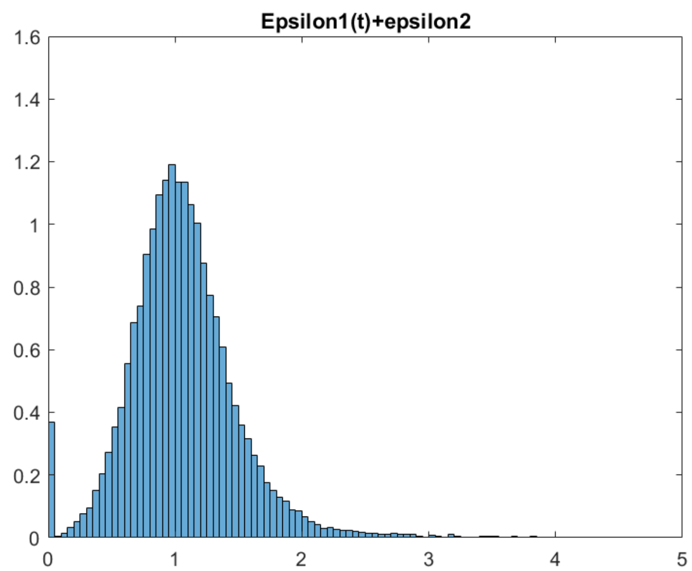

Figure 6: Histogram of ε(t) for Ns.

When the median of the pool is different from zero, we have access to values of ε(t) by taking the ratio of the counts in stations $Y_i(t)$ to the mean of the pool $\mu_s(t)$ (as an estimator of s(t)). The histogram of ε (t) is represented in Fig. 6.

On longer timescales (typically with a moving average of 54 days) we can estimate the drift component as $\mu_2(t)$ (the mean of $\varepsilon_2(i,t)$).

Now that we have the noise model, we have to describe it statistically speaking. Thus, instead of defining error bars, we have the complete distribution of the values.

The probability distribution function (PDF) of $\varepsilon_3(t)$ may be described by a ZA-mixture of t-Location scale (t-LS), for Ns and Ng. They appear in Fig. 7. This distribution is a generalization of the Student distribution: it allows the modeling of a random variable with asymmetry and heavier tails than the normal distribution (Taylor & Verbyla, 2004; Evans, Hastings and Peacock, 2000). The visual closeness between the histogram and the fitted distribution was used as a criteria to select the best PDF among few intuitive candidates.

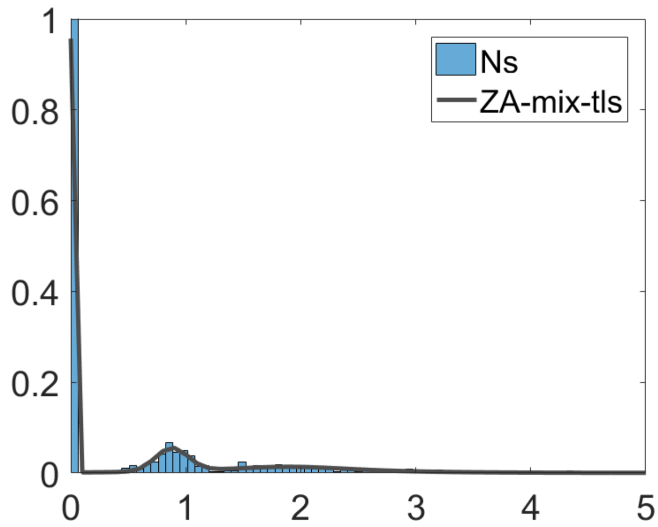

Figure 7: Histogram of ε3(t) for the counts of spots Ns (left) and groups Ng (right). The continuous line shows the fit using a ZA-mixture of tLS distributions. The binning (bw= 0.0917) is the same for both figures and is computed using the Scott procedure for Ns.

In Fig. 7, we immediately notice the important excess of 'true' zeros, together with dominant peaks around one and two for both Ns and Ng. These small excesses at zero represent the most represented value at ow regime: no spots.

Our proxy for s(t) might be equal to zero even if some spots appear several minutes on the sun, due to the non-simultaneity of the observations between stations. It is indeed possible that only part of the network (observing the sun at a different time than other stations) has actually seen these spots. The exact correspondence between Ns and Ng is very interesting: it means that what we observe at that regime is simply groups of only one spot. Even if there are two groups, there are only two spots, i. e. one spot per group. This is exactly what we would expect at low regime of solar activity.

Task 2.3 USET Documentation

The inventory of all the historical information available in ROB logbooks, databases and annual reports started during year 2 is complete. Note that some information is still missing because it is located somewhere else in Europe, Zurich mostly, at the ETH Library. Retrieving it is costly in terms of money and time, and as USET has no specific funding for the curation of its own data, it will be done along with operational tasks, as time permits.

Recommendation: The USET collection is invaluable insofar as it contains all the keys to link past observations to future observations of the Sun (it observes through drawings realized by hand and also through 3 different wavelengths via CCDs), and also to link the past International Sunspot Number to its future. Although the observing station that enables this “link” to survive through time presents minimal costs compared to expensive space missions, ground-based observations of the Sun are never at the center of Belgian or European scientific calls.
The international sunspot number is key to the understanding of the evolution of the Sun’s radiation input on the Earth atmosphere, and thus on the understanding of global warming and its processes. This means that the maintenance of the associated collections and observations should be a priority when decisions are made at the highest levels. And as cost is always important, the ground-based option should be considered extensively, especially structures that already exist.

Task 2.4 USET Quality control

A first iteration of the quality control was achieved between years 2 and 3 enabling the corresponding data to be used in Aparicio et al. (2018). Thanks to the development of DIGISUN2.0 by colleague S. Bechet, a very similar catalogue of sunspot parameters is available from the main station of the WDC-SILSO network: the Locarno station. It extends only from January 2019 to today, but enables a precise analysis of the quality of the data.

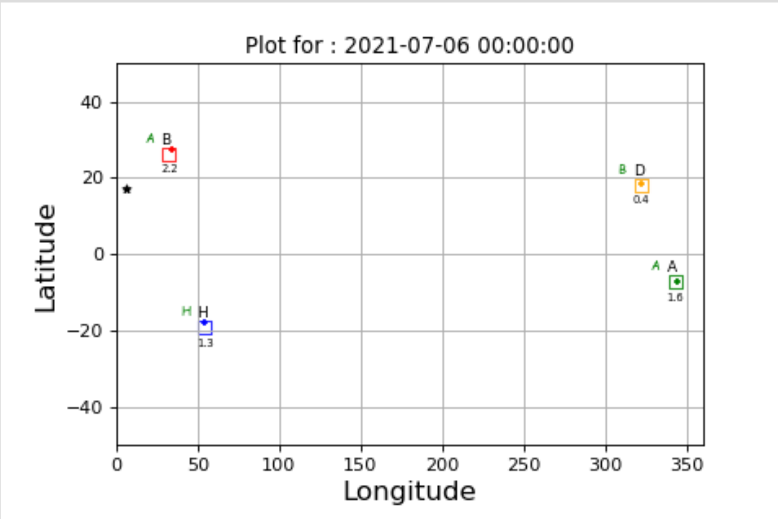

Figure 8: Positions of USET and matched Locarno groups. Matched groups are in the same color, where squares represent group positions from USET and dots represents group positions from Locarno. Group classification from USET group is denoted in black and that of Locarno is marked in cursive green. The distance (Euclidean combination of latitude and longitude) of the group positions are marked below each group. The black star represents any extra group reported by Locarno but missed by USET.

Table IV presents a summary of the content of both catalogs.

Figure 8 shows a comparison for July 6th 2021 of what appears in the USET catalogue with the data from the Locarno station.

We match each group from both catalogues thanks to their positions in heliographic latitude and longitude in degrees, with an average accuracy of 0.3 degrees or less (0.2 in latitude and 0.3 in longitude). Once matched, we can assess the accuracy of the other measurements : in this case, we focus on the number of spots inside groups, the area of the groups and the classification of the groups.

Table IV: Summary of the contents of the USET and Locarno catalogues of sunspot groups

Observatory	USET (complete)	USET	Locarno
Total Days	29813	1006	1006
Start	3/5/1940	1/1/2019	1/1/2019
End	10/19/2021	10/2/2021	10/2/2021
Days observed	20650	765 (76% coverage)	794 (79%)
Zero days	4252	407	422
Non-zero days	16398	358	372
Unobserved Days	9163	233	211
No.of groups	99148	628	681
Matched Groups	455	455	455
Matched days	271	271	271
Unmatched groups	98693	173	226

The accuracy of the matching is presented here in figure 9: it shows the distribution of the distance of matched groups in latitude and longitude. The mode and mean of the distributions are shown as dashed and plain lines respectively.

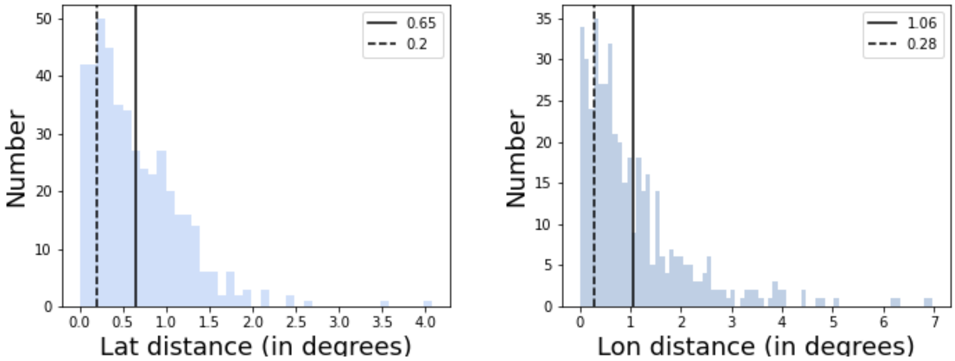

Figure 9: Distribution of the distance in latitude (left) and longitude (right) of matched groups between USET and Locarno catalogues.

As mentioned above the accuracy is within 0.3 degrees, but what does it mean in terms of quality of the data, as quality control of the USET database is the goal of this work? As the cumulative error between the two catalogues is linked to the individual error of both catalogues, it simply means that the error on both catalogues is strictly less than 0.3 degrees.

Now let us compare other parameters from the USET database. Figures 10 to 12 present the matching accuracy in area, number of spots inside a group and classification.

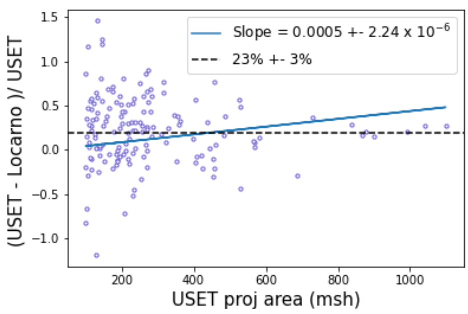

Figure 10 : Relative difference in area (millionth of the solar hemisphere) between the matched groups in USET and Locarno.

The differences between the measured areas from several sources can be explained by different factors. Personal bias from the observers can cause systematic deviations, different locations (same time) can differ because of seing effects (Fligge & Solanki 1997) and differences in observing time can also cause different estimations. Differences in the data reduction methods can cause large deviations in the resulting areas (Pettauer & Brandt 1997).

However, on longer time-scales when data are gathered from several observers in the case of different seeing conditions, there are only two things that mainly influence the systematic differences in addition to the average seeing conditions: these are the observing techniques and the measuring techniques. In the case of Uccle and Locarno, the observing techniques are quite similar (similar setup and instruments) and the measuring techniques are identical insofar as the drawings are measured thanks to DIGISUN 2.0. However, the observers are making the drawings with their own methods. For example, Locarno observers represent the penumbral area around the umbra as a shaded area, while USET observers only draw the contour of the penumbra around the umbra.

The areas of matched groups correspond to within 20% in terms of relative error between the two catalogues, compared to the 3 to 40% difference between several catalogues reported in (Baranyi, et al., 2001, Balmaceda et al. , 2009), it seems to be relatively good. USET seems to report systematically higher areas, a difference that increases with the group size, but that behaviour is to be expected, as the observers have different ways of drawing and both observing setups are similar but different.

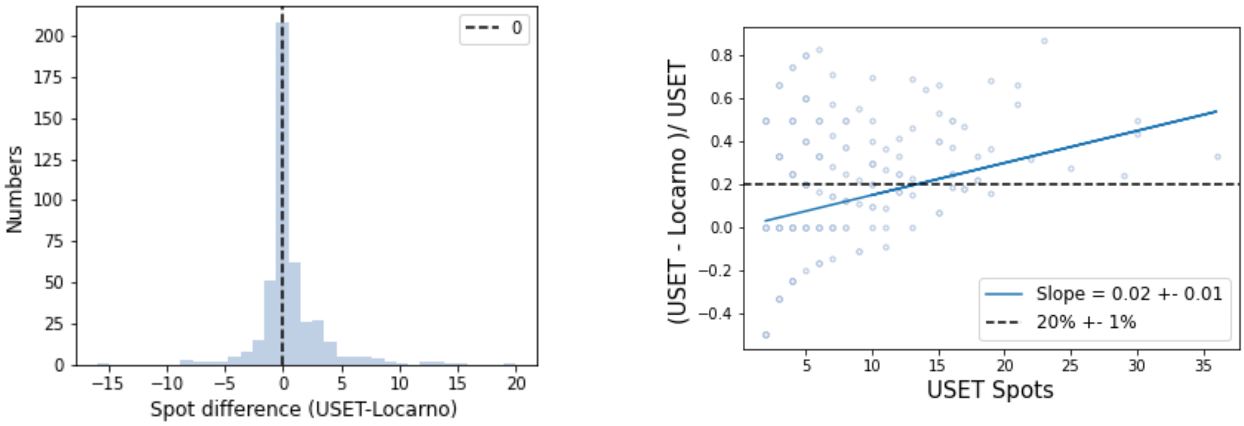

Figure 11 : Difference in number of spots in groups between the matched groups in USET and Locarno (left) and and the relative error between the two measurements (right).

The number of spots is perfectly matched for most of the groups as indicated by the left panel of fig. 11. The difference will increase with the size of groups (the error on counts increases as the square root of the counts). The number of spots in a group correspond to within 20% between both catalogues, which is also a very accurate correspondance. The USET catalogue seems to have a systematically higher number of spots and this difference increases with the size of groups, which is to be expected. Moreover, this systematically higher number of groups explains the systematically higher areas reported by USET. It is simply a different observing and drawing technique. Like the for areas of groups, the differences are linked to the observers bias, and to the seing conditions of the observing location. Long term effects are absent from this comparison as we only compare 3 years of data, but they are often linked to long term evolution of the transparency of the atmosphere (pollution, climate change) and do not affect our dataset significantly.

The last figure (Fig. 12) represents a one-to-one comparison of the Zurich classification of the groups matched between the two catalogues. You can see that the most common occurrences happen on the diagonal of the matrix (similarity matrix), which means both catalogues classify the same way. You can also see that E and F groups are almost absent from the sample. One notable difference though, is the D groups: groups identified as D by the Locarno observers are seen as D or C groups by USET observers, while groups identified by USET observers are seen as D, C or B by Locarno observers.

B groups are bipolar groups with no penumbra on either side, while C groups are bipolar where only one side presents a penumbra and D groups are bipolar with penumbra on both sides. While the definition seems straightforward and should give a clear delimitation of the classes, the identification of the absence or presence of a penumbra around spots is very sensitive to the observers eyesight or to the seing conditions. It seems Locarno has a tendency to see less penumbrae than USET on average. It terms of accuracy or quality assessment, it is difficult to conclude as to which staten has the best practices.

In any case, the Zurich classification scheme itself is well-known to leave too much room for interpretation (Carrasco et al. , 2015) and represents a large part of the inaccuracy of this parameter.

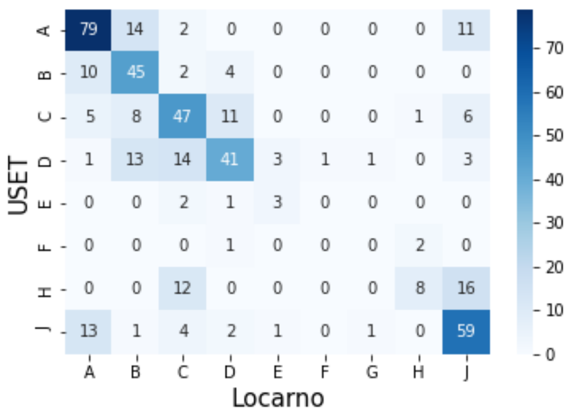

Figure 12 : similarity matrix of the groups classification between USET and Locarno matched groups

Conclusions

Comparing the main USET parameters to the most similar catalogue gives us a very good accuracy that reflects mainly discrepancies outside the setup, such as the observers bias. The next step in this quality control is the comparison with catalogues extracted from different setups (instrumentation), or data support (images or CCD images) and extracted in different ways. We have different catalogues that will enable this comparison on a longer timescale, but of course, as the methods of observing and/or extracting are sometimes widely different, the relative differences will give us less information on the intrinsic error from the method, and more information on extrinsic errors. The analysis presented here, and further comparisons with several other sources will be presented in a series of article to be submitted at the beginning of 2022 that will refer to VALUSUN in the acknowledgements.

WP3

Task 3.1: DIGISUN area computation

Sunspot drawings are the base material available for deriving detailed information about the longterm evolution of the solar cycle. In this context, it is important to have the appropriate tools to convert drawings into exploitable scientific data and to extract as much information as possible from these drawings. While such a tool is obviously important for every observatory with historical sunspot drawings and/or still does daily drawing, there was no generic tool available in the solar physics community. As a side effect, there is a lack of homogeneity among the data and some sunspot drawings collections stay buried in archives due to a lack of manpower to write the appropriate software.

At ROB, there is a collection of 70 years of sunspot drawings, that is still growing with the addition of daily input. The scanning and the first analysis of sunspot drawings are done with an in-house software called DigiSun. In 2018, a new version was designed and the code was rewritten on the basis of a first version which showed its limitation due to its lack of modularity and flexibility. DigiSun 2.0 is a graphical user interface (GUI) with the following functionalities:

(1) Sunspot drawings display, variable image scale and image overlay (large grid on Figure 1) (2) Calculation of the solar axis angles (P, B, L) via an in-house ephemeris module Correspondence between pixels position and heliographic coordinates on the solar disk (3) Recording of sunspot group parameters (number of spots, splitting, Zurich and McIntosh morphological classification). (4) Measurement of the dipole positions and recording of the asymmetry leading/trailing (5) Total area sunspot group measurement

All the parameters of the groups are then saved in an external database for relevant scientific analysis. In addition, it contains a functionality to perform the scanning of drawings making it a unique end-to-end tool to transform drawings into scientific data.

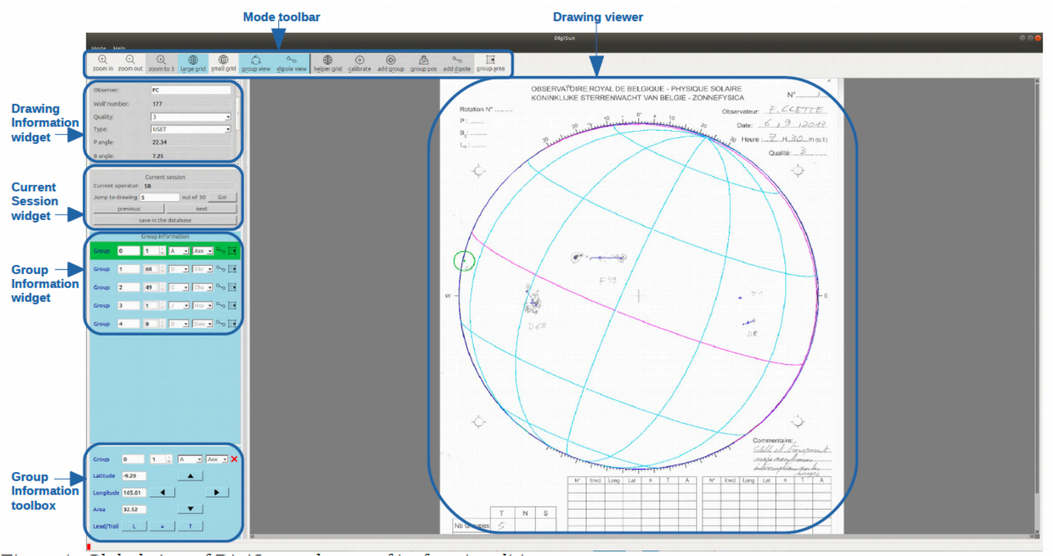

Figure 13 Screenshot of the new DIGISUN 2.0 developed by S. Bechet

Figure 13 above presents a screenshot of the new DIGISUN 2.0 interface. Its advantage compared to the old version is that it is much more user-friendly and thus much faster to use, for the computation of areas and the definition of regions on historical periods, but also for the USET observer on a day-to-day basis.
It is also developed in Python, which enables exportation to every platform in different environments (whether for amateur or professionals). It does not require any extensive coding knowledge to be installed and used.
This software is for all intents and purposes a valorization of the VAL-U-SUN project. Although we were not awarded the valorization grant by BRAIN, we have added value to our project.

Task 3.2 Automatic determination of SILSO quality measurements and reference stations

The long-term analysis of the work as described in task 2.2 is used in the monitoring of the WDC-SILSO stations.

Historically, the observers were rescaled to each other using what we call the k-factors. They were computed yearly and used to fill the gaps in the main observer's values (i.e. the director of the Zurich Observatory). In this study, we need a rescaling for the short-term analysis, of the stations. Thus, we decided to compute an improved version of the k-factors, using ordinary or total least squares or a weighted version of the k-factors. We also use a Kruskal-Wallis test to test on which length of time these rescaling factors have to be computed. This study leads us to an optimized time interval of about 11 months, which is very close to the 1 year used historically.

Of course, for the long term analysis, we want to assess a quantity similar to these k-factors, so the rescaling is not necessary.

We have defined a generic estimator for the long-term error:

$$\widehat{\mu_2}(i,t) = \left(\frac{Y_i(t)}{\hat{\mu}_s(t)}\right)^{\star} \quad \text{when } \hat{\mu}_s(t) > 0$$

where the * operator denotes the smoothing process, namely a moving average of at least 81 days, the ^ refers to a median. We choose $\mu_s(t)$ as the median of the network (all Yi(t) for N stations).

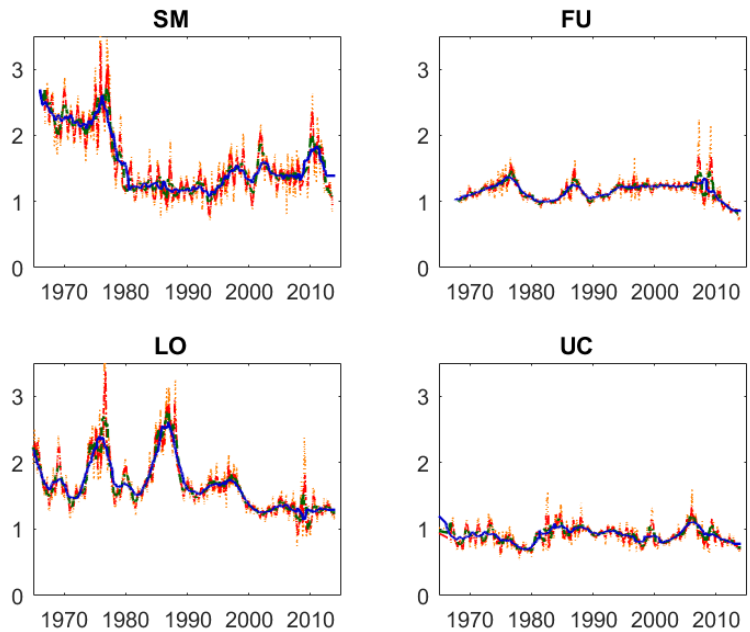

Figure 14: Estimation of the long-term drifts μ2(i,t) of four stations (FU, LO, SM and UC). μ2(i,t) is shown with different MA window lengths: 81 days (orange dotted line), 162 days (red dash-dot line), 1 year (green dashed line) and 2.5years (blue plain line)

Fig. 14 presents the long-term drifts of 4 different stations including our own USET station (code UC). These figures actually show the drift of the pilot station of the network, Locarno (LO) and the stability of USET.

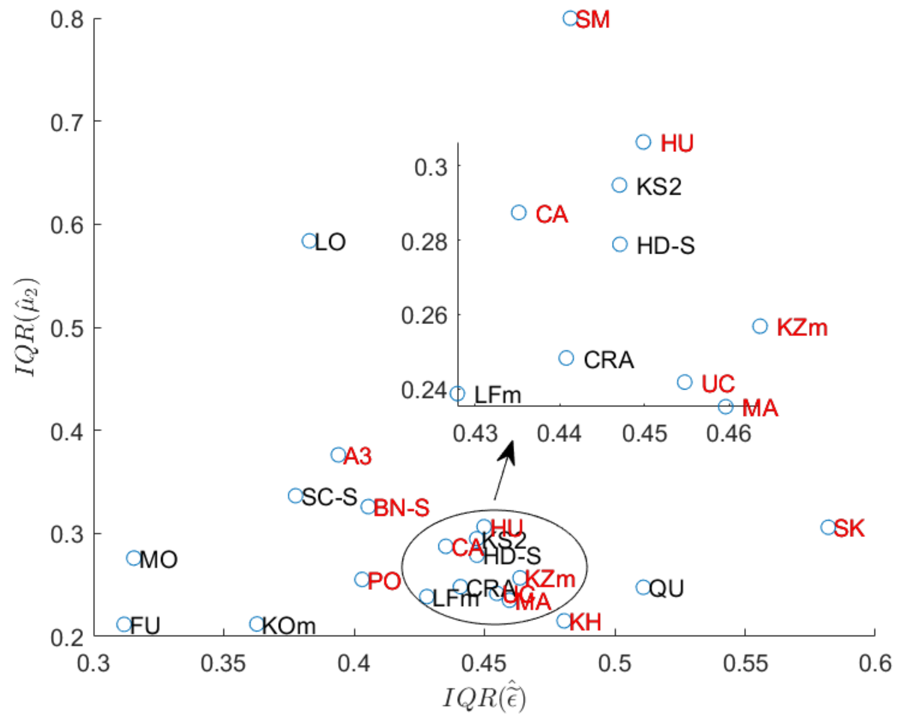

Figure 15. Scatter plot showing the interquartile range of the estimated short-term error ε1(i,t) and the interquartile range of the estimated long-term error μ2(i,t), station by station. Stations in red represent teams of observers while black ones are single observers.

Figure 15 shows a representation of long-term against short-term errors for a sample of stations. It shows the empirical interquartile range, and thus characterizes the stability of the stations outside of minima. In red are the teams of observers: they experience more short-term variability than an individual. For example MO (Mochizuki, Japan), FU and KOm (Koyama, Japan) have low-variability on both the short and long-term. They correspond to long individual observers with stable observation practices. On the other hand, the Locarno (LO) station has a lower long-term stability, while its short-term variability is remarkably low for a professional observatory. As mentioned earlier, this is due to the fact that there is a main observer. Our observing station (USET, UC) shows a large variability in the short-term (due to many observers) but an interesting long-term stability.

Now let us describe in more details the long-term analysis of the work, as it is the part used in the monitoring of the WDC-SILSO stations and also towards the automatic determination of the quality of the SILSO stations.

We propose a nonparametric monitoring that is tailored to the complex features of the sunspot numbers: (a) the missing values, (b) the strong noise, (c) the complex autocorrelation structure and (d) the non-normality. Our method extensively exploits the information contained in the panel to establish a robust IC reference from the network. This allows us to monitor the stations without prior information on their stability. We complete the method by a support vector machine (SVM) procedure that efficiently predicts the size and shape of a shift once an alert has been raised. Although we could manually build a library with typical shapes and sizes to be compared to the deviations, we select the automatic SVM approach instead.

The control scheme is then applied on past observations to study the deviations of the sunspot numbers. The procedure automatically detects major deviations identified recently by hand in some stations. It also unravels many other deviations, unseen in previous analyses. In particular, small and persistent shifts that are difficult to identify manually are detected by the method. The precise information about the deviations predicted by the SVM procedures allows us to determine the causes of some prominent deviations. This sets the ground for a future enhancement of the quality of the series. Moreover, the monitoring procedure provides the possibility to be used in real-time to preserve the long-term stability of the stations.
It also paves the way to a future redefinition of the International Sunspot Number based on several stations that are stable over time.

Step 1: *Pre-processing of the data, estimation of the* ***in-control (IC)*** *parameters, and standardization of all stations (in-control and* ***out-of-control (OC)****) by these parameters.*
We work with the estimation of the long-term error, the mu_2(i,t) defined in the equation above. We first remove the individual level of the stations (subtraction of a moving average of 180 days). Then, we select the IC pool by applying different clustering algorithms on the **mean squared error (MSE)** of each station. We compute the mean and the variance of the IC pool using empirical estimators on a window across the IC stations and along the time. Then, we standardize all stations (IC and OC) by these parameters. We study the correlation of the residuals and fix the length of the studied block to 27 days (cf. step 2).

Step 2 (monitoring): *Design of a control chart using the block bootstrap procedure.*
Here, instead of sampling individual data, we sample blocks of observations with repetitions from the original data. It creates new series of observations with similar correlation as the data. We fix the control limit of the chart using a numerical procedure, to reach a pre-specified rate of false positives.
Then we compare the performance for different types of charts: EWMA, CUSUM, EWMA-GLR, etc. (CUSUM, EWMA, Zhang, 1998, Qiu, 2013). We choose the CUSUM chart, and we adapt it to be robust against missing values.

Step 3: Estimation of shifts sizes and shapes
We use support vector machine (SVM) procedures predict the shifts size and shape on sub-series detected as out-of-control by the CUSUM, for easier problem diagnostic.
The CUSUM gives an alert when a deviation is detected in the data but does not provide information about the characteristics (shape and size) of the shift. Such information is however valuable to assign possible causes to the shift or to adapt the type of alerts that is sent back to the observers. To that end, Cheng et al. (2011) appended a support vector regression (SVR) to the CUSUM. This method is designed to predict, after each alert, the magnitude of shifts in independent and identically normally distributed data that only experience jumps. In the following, we extend Cheng et al. (2011) and design a method that is effective to detect the sizes and the shapes of the deviations in the sunspot number data. This is achieved by a SVM classifier (SVC) (Burges, 1998) in addition to a SVR on top of the chart.

We present here a typical example of major jumps that occurred in the composite Nc smoothed on 27 days. Figure 16 shows the method applied on the data from the station Uccle (UC), in Belgium. UC is stable over time but suffers however from a large jump in 1999, already visible in previous analyses (Clette, 2013). This deviating episode is related to the intense participation of a particular observer that did not count with the same precision as other members of the team. He was recruited at a time where there was a lack of observers but eventually stopped observing after a while.

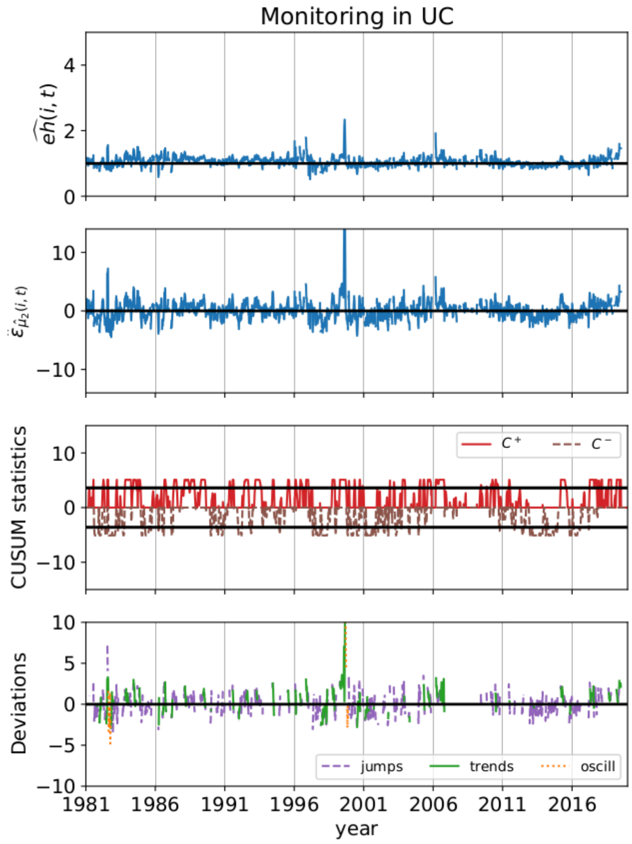

Figure 16: Top to bottom (1) Estimator of the mean of UC. (2) Residuals for Nc smoothed on 27d for Uccle over 1981-2019. (3) (two-sided) CUSUM chart statistics applied on the residuals in square-root scale. Control limits of the chart are represented by the two horizontal thick lines. (4) characteristics of the deviations predicted by the SVR and SVC after each alert.

Depending on which types of events need to be detected, the control parameters have to be adjusted and the gitlab developed by Sophie Mathieu (https://gitlab-as.oma.be/SIDC/SILSO_USET/valusun) contains the necessary explanations to adjust these control parameters.

Recommendations

Recommendations to policymakers

Both collections at the center of the VALUSUN project are, just like all the collections at the center of the BRAIN projects, invaluable.

The USET collection is invaluable insofar as it contains all the keys to link past observations to future observations of the Sun (it observes through drawings made by hand and also through 3 different wavelengths via modern CCDs), and also to link the past International Sunspot Number to its future.

The international sunspot number is key to the understanding of the evolution of the Sun's radiation input on the Earth atmosphere, and thus on the understanding of global warming and its processes. This means that the maintenance of the associated collections and observations should be a priority when decisions are made at the highest levels.

Although the observing station that enables this "link" to survive through time (the Uccle Solar Equatorial Table) presents minimal costs (a few K€/year) compared to space missions (a few M€/year at least), ground-based observations of the Sun are underrepresented in Belgian or European scientific calls. As cost is always important, the ground-based option should be considered extensively, especially structures that already exist. In the context of the Climate crisis we are going through, the images and drawings of the USET station and the SILSO sunspot number are both quantities that are at the base of the radiation budget of the Earth, and as such, these Belgian infrastructures and the associated expertise should be ensured on the longer term, and not subject to short-term financing.

Recommendations to "future" promotors : lessons learned

Contigency plans should be thought about from the start. A project has to be adaptable. I would say the planning can be adapted, as long as the deliverables can be accounted for at the end of the project, or at the time they are needed for serial tasks.

Jobstudents are interesting as manpower goes, but they should be monitored very closely, and the recruitment of job students also has to be pro active as sometimes, no students will spontaneously apply for a job.

Recommendations to BELSPO: future calls ?

The FAIRness of these collections is important, as is stressed by the current BRAIN 2.0 call, but some aspects have been neglected. Preserving an important historical collection that is ongoing involves having the means to continue making the actual observations themselves. An aspect that is rarely taken into account in the calls. This implies hardware, but that is a very small budget, and that is mostly in the Institute's budget, but the manpower to maintain the hardware, and realize the observations is not considered in the long-term.

In summary, the "ongoing" and "long-term" aspects of historical collections are often neglected. For important ongoing collections, 2 to 4 years of funding cannot lead to an efficient functioning point, as the already reduced teams have to answer calls on a regular basis just to maintain a minimum level of activity.

5. DISSEMINATION AND VALORISATION

In order to publicise the VALUSUN project and its results and deliverables, a number of national and international meetings were organised and/or attended (5a and b).

List of VALUSUN meetings

A series of organizational and progress meetings were set up for the project. In this section, we describe the meetings and a summary of the minutes is provided.

- *June 9th 2017: Student reports + meeting for recruitment of a PhD student.*

Present: L. Lefèvre, V. Delouille, C. Ritter, R.von Sachs
Location: ROB
Minutes:
-Review of the reports from the students with C. Ritter (cf. Annexes for a description of the work).
-*Discussion with prospective PhD student Sophie Mathieu.* Questions to assess her knowledge and willingness to learn statistics and solar physics, and her level of computational skills, which are mandatory in both fields. Questions on the context of the PhD project: sunspot numbers, their history and their multiple applications. How does she see herself in this context? Lengthy explanations on the institutes in which she will be working for 4 years.

- *October 16th 2017: Meeting to plan the PhD work.*

Present: L. Lefèvre, O. Lemaître, V. Delouille, C. Ritter, R.von Sachs, S. Mathieu, F. Clette
Location: ROB
Minutes:
Establishment of the specific convention between ROB and UCLouvain for Sophie Mathieu.
Sophie Mathieu is based at UCLouvain, so we establish the planning of visits to ROB for future months.
Account setup at ROB. Set up of Sophie's computer within the ROB network.
Setup of the local and remote access to the SILSO database to work on the stations data directly.

- *December 8th 2017: First official and international meeting of the VAL-U-SUN project.*

Present: L. Lefèvre, O. Lemaître, V. Delouille, C. Ritter, R.von Sachs, S. Mathieu, F. Clette, T. Dudok de Wit, D. Berghmans (director of the Solar Physics department).
Location: ROB
Minutes:
During this meeting we defined the **first version of the model of the Sunspot Number**.
We discussed a first noise model but concluded that it was not appropriate because the term that is meant to model the drift of a station is typically multiplicative and not additive. (A station will count e.g. 10% more sunspots, another 5% less, than a reference, and not 10 sunspots more, or 5 sunspots less).
We note that a similar setting was studied in the case of noise analysis of EUV images. In the case of EUV images, it is possible to graphically represent the relationship between variance and mean by representing (spatially) local variance versus local mean. From such a graph, one may characterize the data and guess appropriate distributions for the error terms.

We assume that the short-term observation error is dominant in the model for time scales smaller than 27 days. The long-term drift in sunspot counts that is particular to a station is supposed to be Gaussian with a specific mean. This term is dominant for periods > 27days.

Problems raised Y1:
How to define the parameter k (scaling between the different stations) properly? Treatment of zero values? How to treat multiplicative error without taking the logarithm for example? How to separate the model in two regimes?

- *March 30th 2018: Progress Meeting.*

Present: L. Lefèvre, V. Delouille, R.von Sachs, S. Mathieu
Location: ROB
Minutes:
- Preparation of the Article for the IEEE meeting in June.
-*Noise Model discussions*.

- *May 8th 2018: Progress meeting.*

Present: L. Lefèvre, C. Ritter, R.von Sachs, S. Mathieu
Location: LLN
Minutes:
This meeting concentrates on the short-term model. We try to correct our model for the presentation of the poster on June 2018 and not spend too much time on new ideas.
Plan for the meeting.
1. Presentation of recent analysis on the conditional distribution (Update by C. Ritter).
2. Discussion on the short-term model (comments by V. Delouille).
3. Discussion on the referee remarks.

- *June 21st 2018: International meeting of the VAL-U-SUN project.*

Present: L. Lefèvre, O. Lemaître, V. Delouille, C. Ritter, R.von Sachs, S. Mathieu, F. Clette, T. Dudok de Wit.
Location: ROB
Minutes:
Choice of the scaling factors to rescale all observers/observing stations to the same observer (discussion with F. Clette). Proxy used for the solar signal (mean, median, more complex?). Choice of the moving average for the mid-term analysis. Discussions with T. Dudok de Wit on the distribution of the values of the different stations when the network gives a signal level s.

Problems raised Y2:
Start thinking about the monitoring procedures.

- *October 3rd 2018: Progress meeting.*

Present: L. Lefèvre, V. Delouille, C. Ritter, R.von Sachs, S. Mathieu
Location: ROB
Minutes:
We decide to write two articles. The first one will contain the model, the distribution of the different terms and the scaling procedure. The second will focus on the monitoring. For the first paper, we aim to publish it in the Astrophysical Journal (https://iopscience.iop.org/journal/0004-637X), the second one will be submitted to a statistical review (to be determined).
Discussions about the choice of the time-scales for the scaling of the different observers to the network.

- *December 3rd 2018: Progress meeting.*

Present: L. Lefèvre, V. Delouille, C. Ritter, R.von Sachs, S. Mathieu
Location: LLN
Minutes:
Discussions about the conditional model of our data. Ordering of the article. Discussions about the estimator of the solar signal: either the median of the network or the filtered median of the network. (solution, create a simulation and check the distribution with, without filtering.). Discrete nature of Ns, Ng and Nc.
Composite Nc:We agree that it is very difficult to retrieve the distribution of the composite Nc = 10Ng + Ns from the linear combination of Ng and Ns as these two variables are dependent and have complex mixture distributions. We will analyze the conditional correlation of Ns and Ng when ISN = s Nc = s.

- *January 25th 2019: Progress meeting.*

Present: L. Lefèvre, V. Delouille, C. Ritter, R.von Sachs, S. Mathieu
Location: Skype
Minutes:
Discussions on the writing of the article to be published in the Astrophysical Journal.

- *April 3rd 2019: Progress Meeting.*

Present: L. Lefèvre, V. Delouille, R.von Sachs, S. Mathieu
Location: Skype
Minutes:
Meeting for the article to be published in the Astrophysical Journal.

1. Article content: should we add/remove content?
2. Article length: text could be written more concisely. Optimal size of the paper.
3. Number of figures and tables. Are they all coherent, necessary?
4. Remarks on the paper.
5. Planning for the next meeting(s) regarding the paper.
6. General planning VALUSUN: next meeting(s)

- *June 5th 2019: Progress meeting.*

Present: L. Lefèvre, C. Ritter, R.von Sachs, S. Mathieu, V. Delouille
Location: Skype
Minutes:
This meeting concentrates on the last corrections to be done to the article, as well the preparation of the ISSI meeting (https://www.issibern.ch/teams/sunspotnoser/index.php/meeting-2-aug-2019/) in August 2019.

- *August 2019: International meeting of the VAL-U-SUN project.*

Present: L. Lefèvre, S. Mathieu, F. Clette, T. Dudok de Wit.
Location: ISSI Bern
Minutes:
Choice of the scaling factors to rescale all observers/observing stations to the same observer (discussion with F. Clette). Proxy used for the solar signal (mean, median, more complex?). Choice of the moving average for the mid-term analysis. Discussions with T. Dudok de Wit on the distribution of the values of the different stations when the network gives a signal level s.

Problems raised Y3:
Start thinking about the monitoring procedures.

- *October 23rd 2019: Progress meeting.*

Present: L. Lefèvre, V. Delouille, C. Ritter, R.von Sachs, S. Mathieu
Location: ROB
Minutes:

1. Presentation of SILSO PhD student: her background, her previous work on tilt angles of sunspot groups, the subject of her thesis.
2. Feedback of the conferences from Sophie: The most recurrent comment was to develop an automated algorithm to extract and count the spots and groups. Sophie is working on this algorithm in her (sparse) free time.
3. Questions and methodology of the monitoring
4. The next meeting is fixed on 23rd January 2020.
5. The methodology should be ready and tested on the data before the summer 2020.

- *January 23rd 2020: Progress meeting.*

Present: V. Delouille, R.von Sachs, S. Mathieu
Location: Skype
Minutes:
Discussions on the next steps for the development of the method of monitoring of the stations and the article that will be published. For the article, the method needs to be

generalized so that it can be applied in different domains.

After this date, most meetings took place remotely because of the working conditions imposed by Covid-19.

- *August 6th 2020: Progress meeting.*

Present: L. Lefèvre, V. Delouille, R.von Sachs, S. Mathieu
Location: Skype
Minutes:
Monitoring article: revision.

- *August 28th 2020: Progress meeting.*

Present: L. Lefèvre, V. Delouille, R.von Sachs, S. Mathieu, C. Ritter
Location: Skype
Minutes:
Monitoring article: revision.

- *October 15th 2020: Progress meeting.*

Present: L. Lefèvre, V. Delouille, R.von Sachs, S. Mathieu, C. Ritter
Location: Skype
Minutes:
Final deliverables/products discussions. What products should the UCLouvain deliver to the ROB/SILSO to enable the application of the error determination and monitoring on a day-to-day or quasi real time basis.

- *October 20th 2020: Progress meeting.*

Present: L. Lefèvre, S. Mathieu, F. Clette
Location: ROB
Minutes:
Final deliverables/products discussions. What shape should the products delivered take ? Start of a Gitlab with all developed programs in Python, accompanied by the necessary documentation.

- *June 28th 2021: Progress meeting.*

Present: L. Lefèvre, V. Delouille, R.von Sachs, S. Mathieu, C. Ritter
Location: Skype
Minutes:
Revision of the PhD thesis of Sophie Mathieu. Deliverables VALUSUN, Gitlab.

- *August 30th 2021: Progress meeting.*

Present: L. Lefèvre, V. Delouille, R.von Sachs, S. Mathieu, C. Ritter, C. Legrand (UCLouvain), B. Frenay (UNamur)
Location: Hybrid Skype/UCLouvain
Minutes:
Pre-defense PhD thesis of Sophie Mathieu.

Problems raised Y4:
Definition of the UCLouvain deliverables for WDC SILSO.

Scientific Dissemination (cf. WP4.2)

Scientific Meetings where the VALUSUN project and products were presented

- **June 2017**, Sunspot Workshop, Brussels, oral presentations, organisation of the International Workshop
- **July 2017** Irkutsk Russia, Oral presentation
- **September 2017** Boulder Sunspot Workshop, US, oral presentations

- **January 2018**, ISSI team meeting, Sunspot Workshop, Bern, oral presentation, chair of session
- **April 2018**, Laure Lefèvre, EGU, Vienna, Austria, oral presentation
- **June 2018**, Sophie Mathieu, Veronique Delouille, IEEE DSW, Lausanne, Poster
- **July 2018,** Laure Lefèvre**,** COSPAR, Pasadena, CA, US, oral presentation
- **October 2018**, Sophie Mathieu, RSSB meeting (Royal Statistical Society of Belgium), Domaine des Hautes Fagnes, Poster
- **October-November 2018**, Veronique Delouille, Laure Lefèvre, SDO Science Workshop, Ghent, poster presentation
- **July 2019,** Laure Lefèvre**,** SC7, Orford, Canada, oral presentation
- **July 2019,** Laure Lefèvre**,** IUGG, Montreal, Canada, 3 oral presentations
- **August 2019,** Laure Lefèvre**,** ISSI, Bern, Switzerland, oral presentation
- **August 2019,** Sophie Mathieu**,** ISSI, Bern, Switzerland, oral presentation
- **August 2019,** Sophie Mathieu, GRETSI, Lille, poster
- **September 2019,** Sophie Mathieu, ML-Helio, Amsterdam, poster

- **June 2021**, Sophie Mathieu, SDSS (Symposium on Data Science and Statistics) virtual meeting, poster

- **October 2021,** Sophie Mathieu, RSSB meeting (Royal Statistical Society of Belgium), University of Liège, oral presentation

Meetings with the follow-up committee

Name: Alexei Pevtsov
Function: co-Chair of the international Astronomical Union's (IAU) Inter-division B and E working group on Coordination of Synoptic Observations of the Sun – US National Solar Observatory – University of Oulu (Finland)
Phone: +1 (575)-434-7011
E-mail: pevtsov@email.noao.edu

Name: Edward Cliver
Function: Co-organizer of the Sunspot Workshops
Phone: +1 (303)-735-8088
E-mail: ecliver@nso.edu

Members of the project met with the follow-up committee members approximately once a year, at international meetings organized for joint projects or virtually via a videoconferencing interface. The project members regularly exchange (at least twice a month) by email or phone information on the progress of the project and any help or data that the members of the follow-up committee can contribute to the project. The results and progress of the project were communicated annually to members of the follow-up committee.
Two ISSI (International Space Science Institute) meetings were used in year 1 and 3 and a COSPAR (http://cospar2018.org/cospar-committees/cospar/) meeting in year 2, to communicate with the follow-up committee members. Due to the global health crisis, no meeting was organised during year 4, but numerous exchanges took place by email.

The coordinator forwarded the information and results of the project to the members of the monitoring committee and organised the annual meetings between the members of the VALUSUN project and the members of the follow-up committee.

Accessibility of the ROB Sunspot Data Collections (WP 4.1)

All project results are published online under an Open Access policy (CC-BY-NC) on the WDC-SILSO and VALUSUN websites (also cf. https://www.astro.oma.be/fr/informations/data-policy/). In this context the database linked to the USET drawings is now accessible at : http://www.sidc.be/valusun/usetdb/ and the WDC-SILSO database is available at http://www.sidc.be/valusun/silsodb/.

Valorization of the VALUSUN work (WP 4.3)

An online consultation platform has been created (cf. 4. Scientific results and recommendation). Through our network of communications (general public: Facebook, Twitter etc…), the public was made aware of the existence of the project and directed to help scientific work towards the consultation platform. Through this platform, they were made aware of the problematics of old and modern counting and the human bias.

Communication channels were also established with more specific communities of users: the observers of the network and the people who use the data published by WDC-SILSO on a monthly basis were made aware of the changes that will soon take place through the WOLF interface (http://www.sidc.be/valusun/wolf/ reference to the website and its use) where the observers enter their data and through the bulletins published on a monthly basis. The software developed by S. Mathieu (UCLouvain, PhD student from VALUSUN) produces outputs regarding the quality of the stations. It is being adapted by the WDC-SILSO team to produce "Quality Issue Notification" e-mail messages to SILSO observers. The team also works on producing, based on the USET database"Sunspot Configuration Bulletin" e-mail messages and send them to relevant space weather consumers. (tasks from WP4)

6. PUBLICATIONS

- Peer reviewed: Aparicio, A.J.P., Lefèvre, L., Gallego, M.C., Vaquero, J.M., Clette, F., Bravo-Paredes, N., Galaviz, P., and Bautista, M.L.: 2018, *Solar Physics*, **293**, 164.
- Peer reviewed: Mathieu, S., von Sachs, R., Ritter, C., Delouille, V., and Lefèvre, L.: 2019, *The Astrophysical Journal*, **886**, 7, https://iopscience.iop.org/article/10.3847/1538-4357/ab4990/pdf
- Non peer-reviewed: Poster and article http://spacescience.org/workshops/mlconference2021/AllAbstracts_F.pdf
- Peer reviewed: submitted to JQT : S. Mathieu, L. Lefèvre, R. von Sachs, V. Delouille, C. Ritter, and F. Clette. Nonparametric monitoring of sunspot number observations. Journal of Quality Technology, 2021. Tentatively accepted.
- S. Mathieu PhD Thesis: https://wwwbis.sidc.be/valusun/web/pdf/thesis.pdf

7. ACKNOWLEDGEMENTS

L. Lefèvre would like to thank all members of the VALUSUN team for their relentless work, including a lot of adaptation done during the COVID 19 pandemics. All members of the VALUSUN team would like to thank the follow-up committee members for their valuable input that helped the project progress in a timely fashion and in the right direction.

The WDC-SILSO team and the VALUSUN team would like to extend their gratitude to all the observers of the network, and particularly those who helped us analyse the results of the monitoring of their stations.

All VALUSUN members, and especially the ones of UCLouvain would like to thank the ISBA for its help during this interesting journey into the Sunspot Numbers. The ROB members would like to thank their UCLouvain partners for hosting meetings and the UCLouvain partners also thank their partner for hosting meetings at the ROB.

The VALUSUN team would also like to warmly thank all the people who have participated to the VALUSUN consultation.

REFERENCES:

- Aparicio, A.J.P., Lefèvre, L., Gallego, M.C., Vaquero, J.M., Clette, F., Bravo-Paredes, N., Galaviz, P., and Bautista, M.L.: 2018, *Solar Physics*, **293**, 164.
- Balmaceda, L.A., Solanki, S.K., Krivova, N.A., and Foster, S.: 2009, {ournal of Geophysical Research (Space Physics) 114, A07104. doi:10.1029/2009JA014299.
- Baranyi, T., Gyori, L., Ludmány, A., & Coffey, H. E. 2001, MNRAS, 323, 223
- Carrasco, V.M.S., Lefèvre, L., Vaquero, J.M., Gallego, M.C.: 2015, Solar Phys. 290, 1445, DOI: 10.1007/s11207-015-0679-y
- Chang and Oh. Does correction factor vary with solar cycle? Journal of Astronomy and Space Sciences, 29(2):97–101, June 2012. doi: 10.5140/JASS.2012.29.2.097.
- Chatzistergos, T., Usoskin, I.G., Kovaltsov, G.A., Krivova, N.A., and Solanki, S.K.: 2017, {\it Astronomy and Astrophysics} {\bf 602}, A69. doi:10.1051/0004-6361/201630045.
- Clette, F., Svalgaard, L., Cliver, E.~W., et al. 2015, IAU General Assembly
- Clette, F., Cliver, E.~W., Lefèvre, L., et al.2016, Solphys, 291, 2479. doi:10.1007/s11207-016-1017-8
- Cliver, E.W., Clette, F., Svalgaard L., and Vaquero, J.M. (2015) , Recalibrating the Sunspot Number (SN): The 3rd and 4th SN Workshops, Cent. Eur. Astrophys. Bull. 39 (2015) 1, 1–19 Dudok de Wit, T., Lefèvre, L., & Clette, F. 2016, *Solar Physics*, **291**, 2709
- M. Evans, N. Hastings, and B. Peacock, Statistical Distributions, John Wiley and Sons, 3 edition, 2000
- Fligge, M., Solanki, S.K., 1997, Solar Physics, 173, 427
- Foster, G. 1999, JAVSO, 27, 177
- D. H. Hathaway. The solar cycle. Living Reviews in Solar Physics, 7:1, March 2010. doi: 10.12942/lrsp-2010-1.
- Hoyt, D. V., Schatten, K. H., & Nesmes-Ribes, E.: 1994, Geophys. Res. Lett. 21, 2067.
- Hoyt, D.V., & Schatten, K.H.: 1998a, Solar Phys. 179, 189.
- Hoyt, D.V., & Schatten, K.H.: 1998b, Solar Phys. 181, 491.
- Kundu, M.R.: 1965, Solar Radio Astronomy (Interscience: New York).
- Lefèvre, L. and Clette, F.: 2011, Astronomy and Astrophysics, 536, L11. doi:10.1051/0004-6361/201118034.
- Lefèvre, L. and Clette, F.: 2014, Solar Physics, 289, 545. doi:10.1007/s11207-012-0184-5.
- Lefèvre, L., Clette, F., (2015) Survey and Merging of Sunspot Catalogs, Sol. Phys., Volume 289, Issue 2, pp.545-561
- Livingston, W., Penn, M.J., & Svalgaard, L. (2012), Astrophys. J., Lett. 757, L8.
- Mathieu, S., von Sachs, R., Ritter, C., Delouille, V., and Lefèvre, L.: 2019, *The Astrophysical Journal*, **886**, 7, https://iopscience.iop.org/article/10.3847/1538-4357/ab4990/pdf
- S. Mathieu, L. Lefèvre, R. von Sachs, V. Delouille, C. Ritter, and F. Clette. Nonparametric monitoring of sunspot number observations. Journal of Quality Technology, 2021. Tentatively accepted.
- D Montgomery. Introduction to Statistical Quality Control. Wiley, New-York, 5th edition, 2004.
- G.E. Morfill, H. Scheingraber, W. Voges, and C.P. Sonett. Sunspot number variations stochastic or chaotic. In C.P. Sonett, M.S. Giampapa, and Matthews, editors, The Sun in Time, pages 30–58. University of Arizona Press, Tucson, United States, January 1991.
- Page, E.S. Cumulative sum charts. Technometrics, 3(1):1–9, 1961. URL https://www.jstor.org/stable/1266472.
- Pettauer, T., Brandt, P.N., 1997, Solar Physics, 175, 197
- B. Schaefer. Improvements for the AAVSO sunspot number. JAAVSO, 26(1): 47–49, 1997.
- S. H. Schwabe, “Sonnenbeobachtungen im Jahre 1843. Von Herrn Hofrath Schwabe in Dessau,” Astron. Nachr. **21**, 233–236 (1844)
- Stewart & Eggleston. The mathematical characteristics of sunspot variations. II.

Astrophysical Journal, 91:72–82, 1940. doi: 10.1086/144147.

- Stewart & Panofsky. The mathematical characteristics of sunspot variations. Astrophysical Journal, 88:385–407, 1938. doi: 10.1086/143994.
- Svalgaard, L., & Hudson, H.S.(2010), in Understanding a Peculiar Solar Minimum, ASP Conference Series 428, eds., R. Cranmer, J.T. Hoeksema, & J.L. Kohl (Astron. Society Pacific: San Francisco), p. 325.
- Svalgaard, L. (2013), J. Space Weather Space Clim. 3, A24.
- J. Taylor and A. Verbyla. Joint modelling of location and scale parameters of the t distribution.Statistical Modelling, 4(2):91–112, 2004.
- Usoskin, I.G., Mursula, K., and Kovaltsov. G. A. Reconstruction of monthly and yearly group sunspot numbers from sparse daily observations. Solar Physics, 218(1-2):295–305, November 2003. doi: https://doi.org/10.1023/B.SOLA.0000013029.99907.97.
- Usoskin, I.G., Kovaltsov, G.A., Lockwood, M., Mursula, K., Owens, M., Solanki, S.K.: 2016, A New Calibrated Sunspot Group Series Since 1749: Statistics of Active Day Fractions. Solar Phys. 291, 2685
- Usoskin, I., Kovaltsov, G., and Kiviaho, W.: 2021, Solar Physics, 296, 13. doi:10.1007/s11207-020-01750-9.
- Qiu. P and Li Z. On Nonparametric Statistical Process Control of Univariate Processes. Technometrics, 53(4):390–405, 2011. 6.3.4
- Qiu. P. Introduction to Statistical Process Control. Texts in Statistical Science, CRC Press, 1 edition, 2013. 6.2.1, 6.2.2
- P. Qiu and D. Xiang. Univariate Dynamic Screening System : An Approach For Identifying Individuals With Irregular Longitudinal Behavior. Technometrics, 56(2):248–260, 2014. 1, 6.3, 6.3.1, 6.3.4, 6.5.1
- P. Qiu, X. Zi, and C. Zou. Nonparametric Dynamic Curve Monitoring. Technometrics, 2017. 6.3, 6.4
- A. Vigouroux and P. Delache. Sunspot numbers uncertainties and parametric representations of solar activity variations. Solar Physics, 152(1):267–274, 1994
- Willamo, T., Usoskin, I.G., Kovaltsov, G.A.: 2017, Updated sunspot group number reconstruc- tion for 1749-1996 using the active day fraction method. Astron. Astrophys. 601, A109
- Willamo, T., Usoskin, I.G., Kovaltsov, G.A.: 2018, A Test of the Active-Day Fraction Method of Sunspot Group Number Calibration: Dependence on the Level of Solar Activity. Solar Phys. 293, 69
- Zhang N. A Statistical Control Chart for Stationary Process Data. Technometrics, 40(1):24–38, 1998. 6.3.4

ANNEXES

Work by C. Ritter's students
Article 1: Mathieu et al. 2019
Article 2: Mathieu et al. 2021
Poster on USET quality control
Programs developed by PhD student S. Mathieu : https://github.com/sophiano/SunSpot