

VAL-U-SUN

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

Contract BR/165/A3/VAL-U-SUN

SUMMARY

Context

The Sunspot Number is the longest scientific experiment still ongoing and a crucial benchmark to study solar activity, space weather as well as climate change. The Royal Observatory of Belgium (ROB) plays a central role in the continuation of this experiment, as it hosts the World Data Center Sunspot Index and Long-term Solar Observations (WDC-SILSO: <https://wwwbis.sidc.be/silso/>). 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.

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 ROB has played an essential role in the study of the Sunspot Number. 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 that is still ongoing today. 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” 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.

Objectives

The purpose of this project is to valorise two sunspot collections kept at ROB, which, at the beginning of the VALUSUN project were not available online or to the level of contemporary modern scientific standards.

The first collection consists is a series of more than 20.000 digitized sunspot drawings acquired at USET since 1940 and a corresponding database that was unverified, i.e. no complete quality checks could be performed and 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 not made accessible to the wider community. As this dataset extends back to 1940, exploiting such a long-term and complete dataset is extremely important for studies of long-term phenomena such as the Earth Climate, but also 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). 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.

Methodology and Results

The USET collection of drawings has been improved with the areas of the groups thanks to the development of a software that we call DIGISUN 2.0 (for VALUSUN, but with ROB self-financed manpower). In addition to being sufficiently user-friendly for mass exploitation, it is also compatible with multiple platforms. It was installed and supported and the Locarno station (the pilot station of the WDC-SILSO network) and thus provides an excellent base of comparison for the USET database of parameters extracted from the USET drawings. Thanks to the new extraction of a Locarno database of sunspot parameters, we could realize a quality control by comparing both datasets on 2019-2021. The next steps will be to compare USET extracted sunspot parameters with other, more inhomogeneous databases with only partly overlapping sunspot parameters. That study has already started and will be published in the first semester of 2022, in the form of two papers: paper I will present DIGISUN2.0 and paper II will present the extensive quality control of the USET sunspot parameters database.

For the second collection of data (i.e., the Sunspot Numbers from all the WDC-SILSO stations), we developed 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 results of this part of the research are published in Mathieu et al. (2019).

We also construct a non-parametric monitoring procedure based on control charts 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. As a result of the present work, 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. The results have been submitted to the Journal of Quality Technology (JQT: <https://asq.org/quality-resources/pub/journal-of-quality-technology>)

All software developed in the context of VALUSUN by the PhD student S. Mathieu, is available at (<https://github.com/sophiano/SunSpot>). The adaptation to the WDC-SILSO reduction pipeline will be realized before mid-2022 for a more efficient exploitation of the data. All data from the WDC-SILSO database have been checked for errors and inconsistencies, and metadata has been added and checked whenever possible. As a result, the data from USET and WDC-SILSO is now quality controlled and contextualised, an important step towards the FAIRness of data that the SILSO-USET center wants to achieve for all its collections.

Conclusions and recommendations

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. The USET database is now accessible at www.sidc.be/valusun/usetdb/.

Thanks to our partners at UCL and their expertise, 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>). The WDC-SILSO database is now available at <http://www.sidc.be/valusun/silsodb/>.

In parallel to this work, a full quality control of the input data in both databases 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).

In conclusion, we would like to point out that preserving, improving and making the datasets of the WDC-SILSO and the USET station available to everyone is of the utmost importance, especially at

this point in time when we realize that preserving the Earth means understanding the evolution of its Climate. And it is not only in the past, but also in the future as both datasets keep growing with ongoing observations.

These datasets contain all the keys to link past observations to future observations of the Sun (USET observes via drawings and 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 rarely 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 like USET and SILSO that already exist.

Keywords

Sunspot Number, Sunspots, statistics, error bars, monitoring

VAL-U-SUN

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

Contrat BR/165/A3/VAL-U-SUN

RESUME

Contexte

Le nombre de taches solaires est la plus longue expérience scientifique encore en cours et une référence cruciale pour étudier l'activité solaire, la météo spatiale ainsi que le changement climatique. L'Observatoire Royal de Belgique (ORB) joue un rôle central dans la poursuite de cette expérience, car il héberge le World Data Center Sunspot Index et Long-term Solar Observations (WDC-SILSO : <https://wwwbis.sidc.be/silso/>). Ce centre mondial de données a pour objectif de collecter des données solaires, ainsi que de produire et de diffuser le nombre international de taches solaires, qui est utilisé dans environ deux cents publications scientifiques chaque année.

Les taches solaires sont des taches sombres apparaissant en groupes sur la surface solaire. Elles sont la manifestation du magnétisme solaire et sont recensées depuis l'invention du télescope au début du XVIIe siècle. En 1843, Samuel Heinrich Schwabe (Schwabe, 1844) a découvert que les séries chronologiques des comptes de taches solaires présentaient une périodicité d'environ 11 ans. On la connaît aujourd'hui sous le nom de cycle solaire, et il affecte profondément l'ensemble du système solaire. Le champ magnétique solaire contenu dans les taches solaires est la force motrice de la variabilité solaire qui influence quotidiennement l'environnement spatial de la Terre. Au milieu du 19e siècle, le Dr Wolf de l'Observatoire de Zürich a créé un indicateur de l'activité solaire en additionnant le nombre total de taches solaires à 10 fois le nombre total de groupes de taches solaires. Avec ses 400 ans de données, ce « nombre de taches solaires » est la plus ancienne série d'observations de l'activité solaire et s'est avéré remarquablement utile, non seulement pour la physique solaire, mais dans tous les domaines scientifiques. La corrélation avec d'autres indices permet de reconstituer la variabilité solaire influençant le climat terrestre au cours des derniers siècles. Le nombre de taches solaires et le cycle solaire sont également une « pierre de Rosette » pour comprendre les étoiles variables de type solaire. La bibliothèque numérique « The NASA Astrophysics Data System » recueille plus de 70 articles arbitrés par an contenant la phrase exacte « sunspot number » dans le résumé ou dans le titre.

Au fil des siècles, l'ORB a joué un rôle essentiel dans l'étude du nombre de taches solaires. En 1939 et pendant la Seconde Guerre mondiale, l'ORB a lancé un programme d'observation solaire en collaboration avec les successeurs du Dr Wolf à l'Observatoire de Zürich qui est toujours en cours aujourd'hui. Les dessins solaires originaux, obtenus à l'instrument ROB Uccle Solar Equatorial Table (USET), ont été conservés et scannés et forment une série homogène de dessins ininterrompus depuis mars 1940. C'est l'une des plus longues collections d'observation stable du Soleil encore active. En 1981, l'Observatoire de Zürich cessa l'activité du World Data Center et en accord avec l'organisation astronomique internationale « COSPAR », le World Data Center fut transféré à l'ORB. A ce titre, le World Data Center « SILSO » est chargé de collecter les observations et de compter le nombre de taches solaires à partir d'un réseau international de stations d'observation et de produire à partir de ces données le nombre international de taches solaires.

Objectifs

Le but de ce projet est de valoriser deux collections de taches solaires conservées à l'ORB, qui, au début du projet VALUSUN n'étaient pas disponibles en ligne et n'étaient pas non plus au niveau des standards scientifiques contemporains.

La première collection est une série de plus de 20.000 dessins de taches solaires numérisés à USET depuis 1940 et la base de données correspondante qui n'a pas encore été contrôlée, c'est-à-dire qu'aucun contrôle de qualité complet n'a pu être effectué et que la base de données est encore incomplète en termes de paramètres extraits. La numérisation de ces dessins a commencé dans le cadre d'un projet européen mais le manque de financement ultérieur nous a empêché d'ajouter des métadonnées essentielles ainsi que de réaliser un contrôle qualité adéquat. Ainsi, il n'a pas été rendu accessible à l'ensemble de la communauté scientifique. Comme cet ensemble de données remonte à 1940, l'exploitation d'un ensemble de données aussi long et complet est extrêmement important pour les études de phénomènes à long terme tels que le climat terrestre, mais aussi afin d'évaluer la qualité des catalogues solaires parallèles.

La seconde collection, la base de données SILSO, contient les nombres de taches et groupes de taches sur le Soleil observés par un réseau mondial depuis 1981 (plus de 530.000 mesures). Il est utilisé sur base mensuelle pour calculer le nombre international de taches solaires. Un seul contrôle est effectué sur chaque station chaque année, et sa base statistique remonte à la création du Nombre de Wolf dans les années 1850. Une supervision complète et cohérente de cette base de données est importante et comme SILSO est le World Data Center pour la détermination de cet indice, il est de notre responsabilité de faire entrer ce nombre international de taches solaires dans le 21^{ème} siècle en exploitant à son plein potentiel cette base de données existante. À cette fin, un effort international a débuté en 2011 et s'est concentré sur les données du passé, du XIX^{ème} siècle à la fin du XX^{ème} siècle. Malheureusement, l'étude réalisée sur cet ensemble de données spécifique de 1981 à aujourd'hui est restée limitée et les normes modernes ne peuvent être atteintes que par l'utilisation de techniques statistiques plus modernes.

L'objectif à atteindre sur nos deux collections est triple : (1) Amener nos bases de données dans le 21^e siècle en ajoutant des métadonnées essentielles, qu'il s'agisse de paramètres supplémentaires ou de techniques d'évaluation et de contrôle qualité. (2) Utiliser la valeur des paramètres solaires enregistrés dans ces bases de données pour répondre aux questions scientifiques d'aujourd'hui et (3) diffuser les collections et la valeur ajoutée à un ensemble prédéfini de publics.

Méthodologie et résultats

La collection de dessins USET a été enrichie avec les zones des groupes grâce au développement d'un logiciel que nous appelons DIGISUN 2.0 (pour VALUSUN, mais avec la main d'œuvre autofinancée de l'ORB). En plus d'être suffisamment convivial pour une exploitation de masse, il est également compatible avec de multiples plateformes. Il a été installé à la station de Locarno (la station pilote du réseau WDC-SILSO) et constitue ainsi une excellente base de comparaison pour les paramètres extraits de la base de données USET. Grâce à la nouvelle base de données de Locarno, nous avons pu réaliser un contrôle qualité en comparant les deux jeux de données sur 2019-2021. Les prochaines étapes consisteront à comparer les paramètres de taches solaires extraits par USET avec d'autres bases de données plus hétérogènes avec des paramètres de taches solaires ne se chevauchant que partiellement. Cette étude a déjà commencé et sera publiée au premier semestre 2022, sous la forme de deux articles : l'article I présentera DIGISUN2.0 et l'article II présentera le contrôle qualité approfondi de la base de données des paramètres des taches solaires USET.

Pour la deuxième collecte de données (c'est-à-dire les nombres de taches solaires de toutes les stations WDC-SILSO), nous avons développé un modèle d'erreur complet pour les nombres de taches solaires dans un cadre multiplicatif. Le modèle décompose les données en un signal physique, commun à différents observateurs, corrompu par trois types d'erreurs, à court terme, à long terme ainsi que pendant les minima solaires. Nous fournissons une analyse complète des

différents termes du modèle, y compris les ajustements paramétriques de leurs distributions. Ce modèle nous permet d'obtenir des estimateurs plus robustes du nombre de taches solaires et de fournir des erreurs pour ces données à chaque instant. Il met également en évidence les écarts à long terme qui se sont produits dans les séries de plusieurs stations d'observation. Bien que spécialement adapté au nombre de taches solaires, le modèle peut également servir de source d'inspiration pour traiter d'autres jeux de données ayant des propriétés similaires. Les résultats de cette partie de la recherche sont publiés dans Mathieu et al. (2019).

Nous avons également développé une procédure de surveillance non paramétrique basée sur des cartes de contrôle et SVM pour détecter efficacement les déviations du nombre de taches solaires au fil du temps. Le schéma est conçu pour fonctionner avec des processus non normalement distribués et auto-corrélés avec des valeurs manquantes. Il fonctionne à différentes échelles et ne nécessite aucune hypothèse paramétrique sur les données pour fonctionner correctement. Cette méthode nous permet d'identifier automatiquement de nombreux écarts dans le nombre de taches solaires, pour la plupart invisibles dans les analyses précédentes et nous aide à trouver les causes profondes de certains changements importants. À la suite de ce projet, ce schéma de contrôle sera mis en œuvre pour surveiller toutes les stations d'observation impliquées dans le processus de comptage, afin d'éviter l'accumulation future d'écarts importants au fil du temps, tels que ceux observés précédemment. Les résultats ont été soumis au Journal of Quality Technology (JQT : <https://asq.org/quality-resources/pub/journal-of-quality-technology>)

L'ensemble des logiciels développés dans le cadre de VALUSUN par la doctorants S. Mathieu, est disponible sur (<https://github.com/sophiano/SunSpot>). L'adaptation au pipeline de réduction WDC-SILSO sera réalisée avant mi-2022 pour une exploitation plus efficace des données. Toutes les données de la base de données WDC-SILSO ont été vérifiées pour les erreurs et les incohérences, et les métadonnées ont été ajoutées et vérifiées chaque fois que possible. En conséquence, les données de USET et du WDC-SILSO ont désormais été soumises à un contrôle qualité et contextualisées, une étape importante pour rendre les données de SILSO-USET FAIR.

Conclusions et Recommendations

L'équipe USET-WDC-SILSO et les stagiaires embauchés grâce au projet ont mis à jour les bases de données tout au long du projet. Cela comprenait le contrôle de la qualité des données, la collecte et l'amélioration des métadonnées, l'ajout d'un paramètre essentiel à la liste existante : la superficie des groupes enregistrée sur les dessins des taches solaires. À cette fin, le développement d'un nouveau logiciel d'analyse des dessins était nécessaire car le précédent s'est avéré inefficace pour déterminer les zones de taches solaires sur les dessins. La base de données USET est désormais accessible sur www.sidc.be/valusun/usetdb/.

Grâce à nos partenaires de l'UCL et à leur expertise, la base de données SILSO (de 1981 à aujourd'hui) a été largement analysée et utilisée, afin d'amener la production du Nombre international de taches solaires dans le 21ème siècle. Une analyse complète des propriétés statistiques du jeu de données a été réalisée et les résultats sont publiés dans Mathieu et al. (2019). Ensuite, une méthode pour surveiller (détecter toute anomalie dans) les données a été développée et soumise au Journal of Quality Technology (JQT : <https://asq.org/quality-resources/pub/journal-of-quality-technology>), et la doctorante Sophie Mathieu a développé un ensemble complet de programmes en Python afin que le World Data Center SILSO puisse désormais appliquer la méthode de détection en temps réel sur les données (<https://github.com/sophiano/SunSpot>). La base de données WDC-SILSO est désormais disponible sur <http://www.sidc.be/valusun/silsodb/>.

Parallèlement à ce travail, un contrôle qualité complet des données d'entrée dans les deux bases de données a été réalisé par l'équipe WDC-SILSO, les métadonnées correspondantes ont été mises à jour et les données ont été rendues conformes au RGPD (https://ec.europa.eu/info/law/law-topic/data-protection_en).

En conclusion, nous tenons à souligner que la préservation, l'amélioration et la mise à disposition des jeux de données du WDC-SILSO et de la station USET sont de la plus haute importance, surtout

à l'heure où l'on se rend compte que préserver la Terre, c'est comprendre l'évolution de son climat. Et ce n'est pas seulement dans le passé, mais aussi dans l'avenir, car les deux ensembles de données continuent de croître avec les observations en cours.

Ces jeux de données contiennent toutes les clés pour lier les observations passées aux observations futures du Soleil (USET observe via des dessins et des CCD), et également pour lier le nombre international de taches passé à son futur. Bien que la station d'observation qui permet à ce « lien » de survivre dans le temps présente des coûts minimes par rapport aux missions spatiales, les observations au sol du Soleil sont rarement au centre des appels scientifiques belges ou européens. Le nombre international de taches solaires est essentiel à la compréhension de l'évolution du rayonnement solaire sur l'atmosphère terrestre, et donc à la compréhension du réchauffement climatique et de ses processus. Cela signifie que la maintenance des collections et des observations associées doit être une priorité lorsque les décisions sont prises aux plus haut niveau. Et comme un fonctionnement à moindre coût est toujours intéressant, l'option au sol doit être largement envisagée, notamment pour des structures comme USET et SILSO qui existent déjà.

Mots clés

Nombre de taches solaires, taches solaires, statistiques, barres d'erreur, surveillance

VAL-U-SUN

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

Contract BR/165/A3/VAL-U-SUN

SAMENVATTING

Context

Het zonnevlekkengetal is het langste wetenschappelijke experiment dat nog gaande is en een cruciale maatstaf om zonneactiviteit, ruimteweer en klimaatverandering te bestuderen. De Koninklijke Sterrenwacht van België (KSB) speelt een centrale rol in de voortzetting van dit experiment, aangezien het gastheer is van de World Data Center Sunspot Index and Long-term Solar Observations (WDC-SILSO: <https://wwwbis.sidc.be/silso/>). Dit World Data Center heeft tot doel zonnegegevens te verzamelen en het International Sunspot Number te produceren en te verspreiden, dat jaarlijks in ongeveer tweehonderd wetenschappelijke publicaties wordt gebruikt.

Zonnevlekken zijn donkere vlekken die in groepen op het zonneoppervlak verschijnen. Het zijn manifestaties van zonnemagnetisme en worden geteld sinds de uitvinding van de telescoop in het begin van de 17e eeuw. In 1843 ontdekte Samuel Heinrich Schwabe (Schwabe, 1844) dat tijdreeksen van het aantal zonnevlekken een periodiek patroon vertonen van ongeveer 11 jaar. Dit werd bekend als de zonnecyclus, die een diepgaand effect heeft op het hele zonnestelsel. Het magnetische zonneveld ingebed in zonnevlekken is de drijvende kracht achter de zonnevariabiliteit die van dag tot dag de ruimteomgeving van de aarde beïnvloedt. In het midden van de 19e eeuw creëerde Dr. Wolf van het Zürich Observatorium een zonneactiviteitsproxy door het totale aantal zonnevlekken op te tellen met 10 keer het totale aantal zonnevlekkengroepen. Met zijn 400 jaar is dit "zonnevlekkengetal" de langst bestaande waarnemingsreeks van zonneactiviteit en is opmerkelijk nuttig gebleken, niet alleen voor de zonnefysica, maar voor een grote verscheidenheid aan natuurwetenschappen. Kruiscorrelatie met andere indices maakt het mogelijk de zonneaandrijving van het klimaat op aarde in de afgelopen paar eeuwen te reconstrueren. Het zonnevlekgetal en de zonnecyclus zijn ook een "Rosetta-steen" voor het begrijpen van zonneachtige veranderlijke sterren. De digitale bibliotheek "The NASA Astrophysics Data System" verzamelt meer dan 70 gerefereerde artikelen per jaar met de exacte uitdrukking "zonnevleknummer" in de samenvatting of in de titel.

Door de eeuwen heen heeft de ROB een essentiële rol gespeeld in de studie van het zonnevlekkengetal. In 1939 en tijdens de Tweede Wereldoorlog startte ROB in samenwerking met de opvolgers van Dr. Wolf op de sterrenwacht van Zürich een zonneobservatieprogramma dat nog steeds loopt. De originele zonetekeningen, verkregen met het ROB Uccle Solar Equatorial Table (USET)-instrument, zijn bewaard en gescand en vormen een homogene reeks ononderbroken tekeningen sinds maart 1940. Het is een van de langste stabiele observatiecollecties van de zon die nog steeds actief is. In 1981 stopte het Zürich Observatory met de World Data Center-activiteit en in overeenstemming met de internationale astronomische organisatie "COSPAR" werd het World Data Center overgedragen aan ROB. Als zodanig is het ROB World Data Center "SILSO" verantwoordelijk voor het verzamelen van waarnemingen en tellingen van zonnevlekken van een internationaal en wereldwijd netwerk van waarnemingsstations en voor het produceren van deze gegevens het International Sunspot Number.

Doelen

Het doel van dit project is het valoriseren van twee zonnevlekkencollecties die bij ROB worden bewaard en die aan het begin van het VALUSUN-project niet online beschikbaar waren of op het niveau van de hedendaagse moderne wetenschappelijke standaarden.

De eerste collectie bestaat uit een reeks van meer dan 20.000 gedigitaliseerde zonnevlektekeningen die sinds 1940 bij USET zijn verworven en een bijbehorende database die niet geverifieerd was, d.w.z. er konden geen volledige kwaliteitscontroles worden uitgevoerd en de database is nog steeds onvolledig in termen van geëxtraheerde parameters. De digitalisering van deze tekeningen begon in het kader van een Europees project, maar het gebrek aan latere financiering verhinderde ons om essentiële metadata toe te voegen en een goede kwaliteitscontrole te bereiken. Zo werd het niet toegankelijk gemaakt voor de bredere gemeenschap. Aangezien deze dataset teruggaat tot 1940, is het exploiteren van zo'n lange-termijn en volledige dataset uiterst belangrijk voor studies van langetermijnfenomenen zoals het aardklimaat, maar ook om de kwaliteit van parallelle zonnecatalogi te beoordelen.

De tweede collectie, de SILSO-database, bevat de aantallen vlekken en groepen vlekken op de zon die sinds 1981 door een wereldwijd netwerk zijn waargenomen (meer dan 530.000 metingen). Het wordt maandelijks gebruikt om het internationale zonnevlekkengetal te berekenen. Elk jaar wordt er slechts één controle uitgevoerd op elk station en de statistische basis gaat terug tot de oprichting van het Wolf-nummer in de jaren 1850. Een volledig en consistent toezicht op deze database is belangrijk en aangezien SILSO het World Data Center is voor de bepaling van deze index, is het onze verantwoordelijkheid om het geëxtraheerde zonnevlekkengetal in de 21e eeuw te brengen door deze bestaande database optimaal te benutten. Hiertoe is in 2011 een internationale inspanning gestart, gericht op de gegevens uit het verleden, van de 19e eeuw tot het einde van de 20e eeuw. Helaas bleef het onderzoek naar deze specifieke dataset uit 1981 beperkt en kunnen moderne standaarden alleen worden bereikt door het gebruik van modernere statistische technieken.

Het doel dat we met onze twee collecties willen bereiken, is drieledig: (1) Breng onze databases naar het moderne tijdperk door essentiële metadata toe te voegen, of het nu gaat om aanvullende parameters of technieken voor kwaliteitsbeoordeling en kwaliteitscontrole. (2) Gebruik de waarde van zonneparameters die in deze databases zijn vastgelegd om de wetenschappelijke vragen van vandaag aan te pakken en (3) de collecties en toegevoegde waarde te verspreiden onder een vooraf gedefinieerde reeks doelgroepen.

Methodologie en resultaten

De USET-verzameling tekeningen is verbeterd met de gebieden van de groepen dankzij de ontwikkeling van een software die we DIGISUN 2.0 noemen (voor VALUSUN, maar met ROB zelf gefinancierde mankracht). Behalve dat het voldoende gebruiksvriendelijk is voor massa-exploitatie, is het ook compatibel met meerdere platformen. Het werd geïnstalleerd en ondersteund en het Locarno-station (het pilootstation van het WDC-SILSO-netwerk) en biedt dus een uitstekende vergelijkingsbasis voor de USET-database met parameters die zijn geëxtraheerd uit de USET-tekeningen. Dankzij de nieuwe extractie van een Locarno-database van zonnevlekparameters konden we een kwaliteitscontrole realiseren door beide datasets op 2019-2021 te vergelijken. De volgende stappen zullen zijn om USET-geëxtraheerde zonnevlekparameters te vergelijken met andere, meer inhomogene databases met slechts gedeeltelijk overlappende zonnevlekparameters. Die studie is al gestart en zal in het eerste semester van 2022 worden gepubliceerd in de vorm van twee papers: paper I zal DIGISUN2.0 presenteren en paper II zal de uitgebreide kwaliteitscontrole van de USET sunspot parameters database presenteren.

Voor de tweede verzameling van gegevens (d.w.z. de zonnevlekkengetallen van alle WDC-SILSO-stations), ontwikkelden we een uitgebreid foutenmodel voor de zonnevlekkengetallen in een multiplicatief raamwerk. Het model ontleedt de gegevens in een fysiek signaal, gemeenschappelijk voor verschillende waarnemers, beschadigd door drie soorten fouten, zowel op korte, lange termijn als tijdens zonneminima. We bieden een volledige analyse van de verschillende termen van het

model, inclusief parametrische passingen van hun distributies. Dit model stelt ons in staat om robuustere schatters van de zonnevlekkengetallen te verkrijgen en om op elk moment fouten voor die gegevens te verschaffen. Het belicht ook de langetermijnafwijkingen die zich hebben voorgedaan in de afgelopen reeks van verschillende waarnemingsstations. Hoewel speciaal aangepast aan de zonnevlekkengetallen, kan het model ook dienen als inspiratiebron voor het behandelen van andere datasets met vergelijkbare eigenschappen. De resultaten van dit deel van het onderzoek zijn gepubliceerd in Mathieu et al. (2019).

We construeren ook een niet-parametrische monitoringprocedure op basis van controlekaarten en ondersteunen vectormachines om de afwijkingen van zonnevlekkengetallen in de loop van de tijd efficiënt te detecteren. Het schema is ontworpen om te werken met niet-normaal verdeelde en autogecorreleerde processen met mogelijk ontbrekende waarden. Het werkt op verschillende schalen en vereist geen parametrische veronderstelling over de gegevens om correct te werken. Deze methode stelt ons in staat om automatisch veel afwijkingen in de zonnevlekkengetallen te identificeren, meestal ongezien in eerdere analyses, en helpt ons de hoofdoorzaken van enkele opvallende verschuivingen te vinden. Als resultaat van de huidige werkzaamheden zal dit controleschema worden geïmplementeerd om alle waarnemingsstations die betrokken zijn bij het telproces te bewaken, om toekomstige opbouw van grote afwijkingen in de tijd, zoals eerder waargenomen, te voorkomen. De resultaten zijn ingediend bij het Journal of Quality Technology (JQT: <https://asq.org/quality-resources/pub/journal-of-quality-technology>)

Alle software die in het kader van VALUSUN is ontwikkeld door de promovendus S. Mathieu, is beschikbaar op (<https://github.com/sophiano/SunSpot>). De aanpassing aan de WDC-SILSO reductiepijplijn zal voor medio 2022 gerealiseerd worden voor een efficiëntere exploitatie van de data. Alle data uit de WDC-SILSO database zijn gecontroleerd op fouten en inconsistenties, en waar mogelijk zijn metadata toegevoegd en gecontroleerd. Als gevolg hiervan zijn de gegevens van USET en WDC-SILSO nu op kwaliteit gecontroleerd en gecontextualiseerd, een belangrijke stap in de richting van de EERLIJKheid van gegevens die het SILSO-USET-centrum voor al zijn collecties wil bereiken.

Conclusies en Aanbevelingen

Het USET-WDC-SILSO-team en de jobstudenten die dankzij het project werden ingehuurd, hebben de database die bij de tekeningen hoort, gedurende het hele project bijgewerkt. Dit omvatte controle van de gegevenskwaliteit, het verzamelen en verbeteren van metagegevens, de toevoeging van een essentiële parameter aan de bestaande lijst: het gebied van de groepen die op de zonnevlekken tekeningen zijn vastgelegd. Hiertoe was de ontwikkeling van nieuwe software voor het analyseren van tekeningen noodzakelijk, aangezien de vorige inefficiënt bleek om zonnevlekgebieden op tekeningen te bepalen. De USET-databank is nu toegankelijk via www.sidc.be/valusun/usetdb/.

Dankzij onze partners bij UCL en hun expertise, is de SILSO-database (1981-heden) uitgebreid geanalyseerd en gebruikt om de exploitatie ervan voor de productie van het internationale zonnevlekkengetal in de 21e eeuw te brengen. Een volledige analyse van de statistische eigenschappen van de dataset is gerealiseerd en de resultaten zijn gepubliceerd in Mathieu et al. (2019). Vervolgens is er een methode ontwikkeld om eventuele afwijkingen in de gegevens te monitoren (detecteren) en ingediend bij het Journal of Quality Technology (JQT: <https://asq.org/quality-resources/pub/journal-of-quality-technology>), en promovenda Sophie Mathieu ontwikkelde een volledige set programma's in Python zodat het World Data Center SILSO de detectiemethode nu in realtime op de data kan toepassen (<https://github.com/sophiano/SunSpot>). De WDC-SILSO-databank is nu beschikbaar op <http://www.sidc.be/valusun/silsodb/>.

Parallel aan dit werk heeft het WDC-SILSO-team een volledige kwaliteitscontrole van de invoergegevens in beide databases gerealiseerd, zijn de bijbehorende metagegevens bijgewerkt en zijn de gegevens AVG-conform gemaakt (https://ec.europa.eu/info/law/law-topic/data-protection_en).

Tot slot willen we erop wijzen dat het behoud, de verbetering en het beschikbaar stellen van de datasets van de WDC-SILSO en het USET-station voor iedereen van het grootste belang is, vooral in deze tijd waarin we ons realiseren dat het behoud van de aarde begrip betekent de evolutie van zijn klimaat. En het is niet alleen in het verleden, maar ook in de toekomst, aangezien beide datasets blijven groeien met voortdurende observaties.

Deze datasets bevatten alle sleutels om eerdere waarnemingen te koppelen aan toekomstige waarnemingen van de zon (USET neemt waar via tekeningen en CCD's), en ook om het vroegere internationale zonnevlekkengetal te koppelen aan zijn toekomst. Hoewel het waarnemingsstation dat deze "verbinding" in staat stelt om door de tijd te overleven minimale kosten met zich meebrengt in vergelijking met dure ruimtemissies, staan waarnemingen van de zon op de grond zelden centraal in Belgische of Europese wetenschappelijke oproepen.

Het internationale zonnevlekkengetal is essentieel voor het begrip van de evolutie van de stralingsinput van de zon op de aardatmosfeer, en dus voor het begrip van de opwarming van de aarde en haar processen. Dit betekent dat het onderhoud van de bijbehorende collecties en observaties een prioriteit moet zijn bij beslissingen op het hoogste niveau. En aangezien kosten altijd belangrijk zijn, moet de grondgebonden optie uitgebreid worden overwogen, met name structuren zoals USET en SILSO die al bestaan.

Keywords:

Zonnevlekkengetal, Zonnevlekken, statistieken, foutbalken, monitoring