

Northumbria Research Link

Citation: Georgoulis, Manolis K., Bloomfield, Shaun, Piana, Michele, Massone, Anna Maria, Soldati, Marco, Gallagher, Peter T., Pariat, Etienne, Vilmer, Nicole, Buchlin, Eric, Baudin, Frederic, Csillaghy, Andre, Sathiapal, Hanna, Jackson, David R., Alingery, Pablo, Benvenuto, Federico, Campi, Cristina, Florios, Konstantinos, Gontikakis, Constantinos, Guennou, Chloe, Guerra, Jordan A., Kontogiannis, Ioannis, Latorre, Vittorio, Murray, Sophie A., Park, Sung-Hong, von Stachelski, Samuel, Torbica, Aleksandar, Vischi, Dario and Worsfold, Mark (2021) The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era. *Journal of Space Weather and Space Climate*, 11. p. 39. ISSN 2115-7251

Published by: EDP Sciences

URL: <https://doi.org/10.1051/swsc/2021023> <<https://doi.org/10.1051/swsc/2021023>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/48757/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)





















**Northumbria
University**
NEWCASTLE



University**Library**

The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era

Manolis K. Georgoulis^{1,*} , D. Shaun Bloomfield^{2,3} , Michele Piana^{4,5} , Anna Maria Massone^{4,5} , Marco Soldati⁶ , Peter T. Gallagher^{3,7} , Etienne Pariat⁸ , Nicole Vilmer⁸, Eric Buchlin⁹ , Frederic Baudin⁹, Andre Csillaghy⁶, Hanna Sathiapal⁶, David R. Jackson¹⁰ , Pablo Alingery⁹, Federico Benvenuto⁴ , Cristina Campi^{5,11} , Konstantinos Florios^{1,12} , Constantinos Gontikakis¹ , Chloe Guennou⁸ , Jordan A. Guerra^{3,13} , Ioannis Kontogiannis^{1,14} , Vittorio Latorre⁴, Sophie A. Murray^{3,7} , Sung-Hong Park^{3,15} , Samuel von Stachelski⁶, Aleksandar Torbica⁶, Dario Vischi⁶, and Mark Worsfold¹⁰

¹ RCAAM of the Academy of Athens, 11527 Athens, Greece

² Department of Mathematics, Physics & Electrical Engineering, Northumbria University, NE1 8ST Newcastle upon Tyne, UK

³ School of Physics, Trinity College Dublin, Dublin 2, Ireland

⁴ Dipartimento di Matematica, Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy

⁵ CNR – SPIN Genova, Via Dodecaneso 33, 16146 Genova, Italy

⁶ University of Applied Sciences & Arts Northwestern Switzerland, 5210 Windisch, Switzerland

⁷ School of Cosmic Physics, Dublin Institute for Advanced Studies, D02 XF85 Dublin, Ireland

⁸ LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 75014 Paris, France

⁹ Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, 91405, Orsay, France

¹⁰ Met Office, EX1 3QS Exeter, UK

¹¹ Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, 35131 Padova, Italy

¹² Department of Chemical Engineering, National Technical University of Athens, 15772 Zografou, Greece

¹³ Department of Physics, Villanova University, 800 E Lancaster Ave., Villanova, 19085 PA, USA

¹⁴ Leibniz-Institut für Astrophysik Potsdam (AIP), 14482 Potsdam, Germany

¹⁵ Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, 464-8601 Aichi, Japan

Received 28 April 2020 / Accepted 26 May 2021

Abstract – The European Union funded the FLARECAST project, that ran from January 2015 until February 2018. FLARECAST had a research-to-operations (R2O) focus, and accordingly introduced several innovations into the discipline of solar flare forecasting. FLARECAST innovations were: first, the treatment of hundreds of physical properties viewed as promising flare predictors on equal footing, extending multiple previous works; second, the use of fourteen (14) different machine learning techniques, also on equal footing, to optimize the immense Big Data parameter space created by these many predictors; third, the establishment of a robust, three-pronged communication effort oriented toward policy makers, space-weather stakeholders and the wider public. FLARECAST pledged to make all its data, codes and infrastructure openly available worldwide. The combined use of 170+ properties (a total of 209 predictors are now available) in multiple machine-learning algorithms, some of which were designed exclusively for the project, gave rise to changing sets of best-performing predictors for the forecasting of different flaring levels, at least for major flares. At the same time, FLARECAST reaffirmed the importance of rigorous training and testing practices to avoid overly optimistic pre-operational prediction performance. In addition, the project has (a) tested new and revisited physically intuitive flare predictors and (b) provided meaningful clues toward the transition from flares to eruptive flares, namely, events associated with coronal mass ejections (CMEs). These leads, along with the FLARECAST data, algorithms and infrastructure, could help facilitate integrated space-weather forecasting efforts that take steps to avoid effort duplication. In spite

*Corresponding author: manolis.georgoulis@academyofathens.gr

of being one of the most intensive and systematic flare forecasting efforts to-date, FLARECAST has not managed to convincingly lift the barrier of stochasticity in solar flare occurrence and forecasting: solar flare prediction thus remains inherently probabilistic.

Keywords: Sun / solar flares / solar flare forecasting / machine learning / big data / computer science

Ὁ ἥλιος οὐ μόνον... νέος ἐφ' ἡμέρη
ἐστίν, ἀλλ' αἰεὶ νέος συνεχῶς.

The Sun is young every day,
incessantly and eternally.

Ἡράκλειτος, ~500 Π.Κ.Ε.

Heraclitus, ~500 BCE

1 Introduction

The first decades of the 21st century have seen the transformative effect of the ever-increasing, widespread use of wireless technologies. Enhanced by the equally expansive use of the internet, these technologies have claimed, and are expected to continue claiming, a crucial part of our everyday routine, from services to communications and from information to edutainment. Space-based satellite technologies have also been instrumental in establishing wireless capabilities to a degree that few could predict, even by the standards of the late 20th century.

When relying on space, however, it is imperative to keep in mind the decisive effects of our magnetically active star, the Sun. Simultaneously with the expansion of human capabilities came increased awareness of the adverse effects of space weather, namely, the short-term (hours to days) impact of solar magnetic activity, from the fast solar wind spewed by extended coronal holes to the storm-like transport of solar eruptions through the entire heliosphere. Electromagnetic and particulate emission from solar eruptions can cause anything from short-lived, relatively unimpactful disruptions to major damage in satellite infrastructure, on top of the biological hazards they pose to exposed humans in space conditions, either during extravehicular activities or in future manned space travel and missions to Moon and Mars (see, for example, ESA's¹ Moon Village and NASA's Artemis Programs).

Exceptional solar flares and eruptive manifestations, up to the first flare observed by Carrington (1859), are among deep-space phenomena whose repercussions go past the ionosphere, reaching down to aviation altitudes and even to Earth's surface. Figure 1 portrays this impact in an image produced by the University of Applied Sciences and Arts of Northwestern Switzerland (FHNW) partner of the FLARECAST Consortium. There, one sees ramifications spanning from what we already knew before the space age (i.e., the aurora, long-range electrical power networks or radar disruptions) to any applications that GPS or Galileo enable. During the space age we have seen some solar eruptions that have caused major effects in May 1967, March 1989, and October–November 2003, although solar flares associated with these eruptions were arguably smaller than the Carrington flare (Cliver & Dietrich, 2013). However, while cruising on the far-side of the Sun in July 2012, the STEREO-A spacecraft detected the transit of an extreme

solar eruption that was characterized as a Carrington-scale event (Baker et al., 2013). Geospace was spared from that eruption but, statistically, in future events it will not be (see, e.g., Riley & Love, 2017).

The staggering short- and long-term financial impact of extreme space weather has been delineated in a series of recent works (e.g., MacAlester & Murtagh, 2014; Oughton et al., 2017; Eastwood et al., 2018), as well as in governmental guidelines and action plans, such as the *US National Space Weather Strategy and Action Plan* (2015, 2019²) and the *UK Space Weather Preparedness Strategy* (2015³). Of particular interest, however, is the 2008 National Research Council's *Severe Space Weather Events: Understanding Societal and Economic Impacts Workshop Report* (NRC, 2008) that portrays in its Figure 3.1 the nonlinear inner workings and interconnections of sectors comprising our societal fabric. Like domino bricks, if any of these sectors fails due to extreme space weather, the ramifications will be hard to imagine and even harder to mitigate. The reliable forecasting of extreme space weather, therefore, upgrades to a major challenge of our times.

Energetic events accompanying major solar eruptions are by far the main agents of extreme space weather. These events comprise three distinct aspects: solar flares, CMEs, and SEP events. A reliable forecasting, therefore, should encompass three very different forecast efforts with unique characteristics and challenges. In solar flare prediction, that is the topic of discussion in this work, there is no early warning for the flares' X- and γ -ray photons. Only a slim window of ~10–12 min exists for flare-accelerated SEPs, if any (Haggerty & Roelof, 2002; Rust et al., 2008). To address the lack of advance warning, therefore, major solar flares – and flare-related, impulsive SEP events, by extension – need to be predicted well before their occurrence (i.e., several hours to 1–2 days in advance). There are currently significant shortcomings in our flare forecasting ability, as sections below will show. In addition, CMEs, particularly the fastest ones that are stemming from solar ARs and are virtually associated one-to-one to major flares (Yashiro et al., 2005, but see Liu et al., 2016 for an exceptional active region), should ideally be predicted along with flares and SEPs (e.g., Anastasiadis et al., 2017). There is a window of inner-heliospheric transit ranging between ~20 h and 2–3 days after the initial, near-Sun detection of CMEs until they reach

¹ All abbreviations and acronyms used hereafter are explained in Appendix D.

² Available at <https://trumpwhitehouse.archives.gov/wp-content/uploads/2019/03/National-Space-Weather-Strategy-and-Action-Plan-2019.pdf>.

³ Available at <https://www.gov.uk/government/publications/space-weather-preparedness-strategy>.

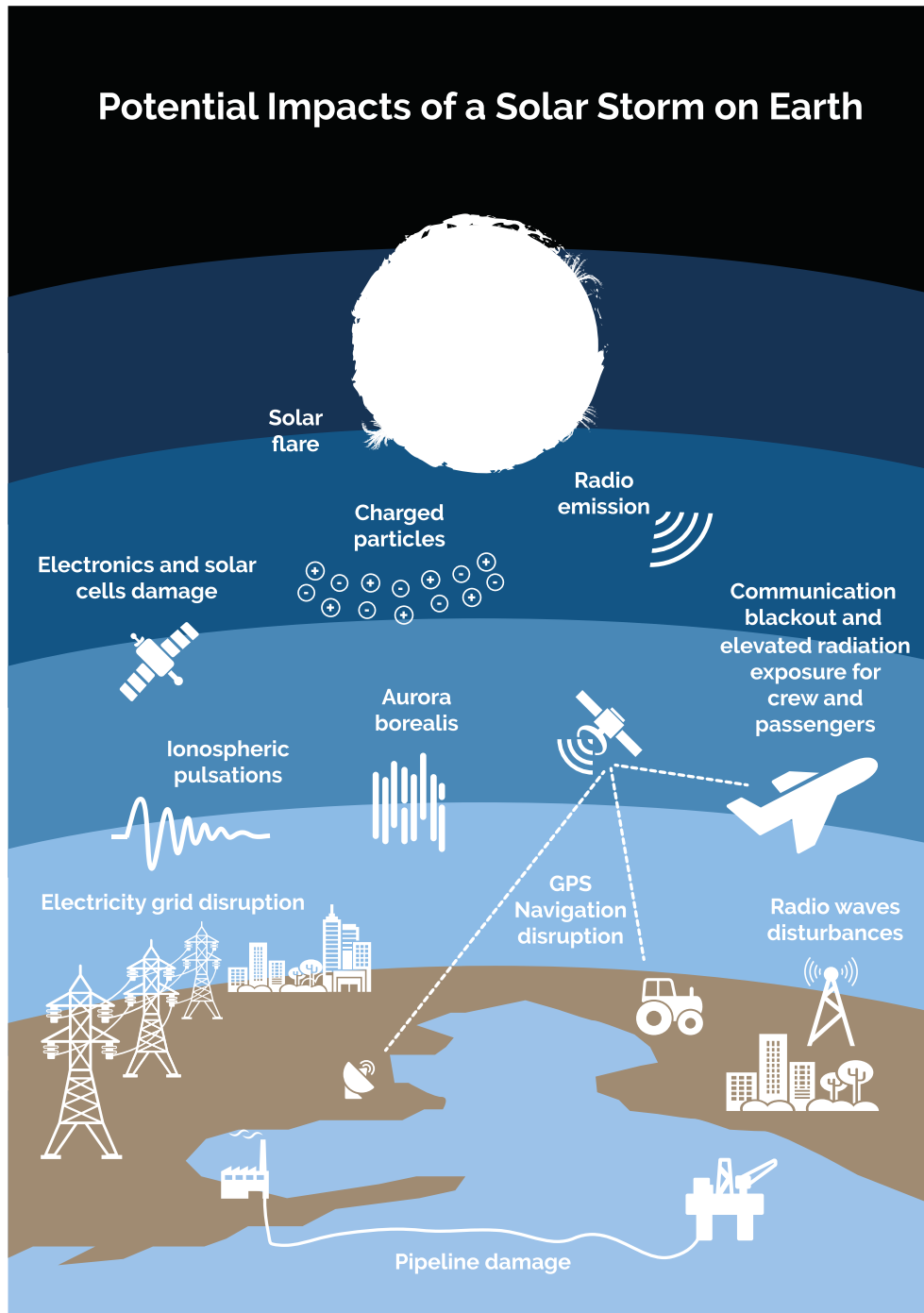


Fig. 1. A graphical representation of extreme space weather and its effects on contemporary technology and infrastructure. Credit: FHNW.

geospace. If Earth-directed, their arrival time and geoeffectiveness (i.e., their potential ability to trigger a geomagnetic storm) should also be predicted in advance (e.g., Möstl et al., 2014; Mays et al., 2015). Finally, CME-shock-accelerated SEP events may arrive at geospace hours after the source solar eruption or, in the worst-case scenario, even before the CME registers in near-Sun height-time diagrams (Reames, 2017; Malandraki & Crosby, 2018, and references therein). We also need to know in advance the temporal profile of the SEP event and its peak flux or fluence, particularly for proton energy channels

exceeding 50–100 MeV, as per NOAA guidelines and recently defined benchmarks⁴.

Although an ultimate goal, we still seem to be far from achieving an integrated platform for the prediction of all extreme space weather manifestations. Among them, solar flare prediction has historically been humanity’s first stride. Since the

⁴ See the US Space Weather Phase 1 Benchmarks at <https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/06/Space-Weather-Phase-1-Benchmarks-Report.pdf>.

1980s, there have been persistent efforts toward flare prediction introducing a wealth of physical, semi-empirical or empirical AR properties and proxies that have been claimed to hold a flare-predictive capability. A short, non-exhaustive review of these properties, including an effort to group them into different categories, appears in Georgoulis (2012).

However, earlier efforts (e.g., Leka & Barnes, 2003b, 2007; Barnes & Leka, 2008) aiming to assess the relative performance of these properties indicated that, on one hand, none was solely capable to predict flares reliably while, on the other hand, when a capability to simultaneously test multiple properties was achieved, the predictive information contained in the full property set was highly redundant. The first comparative evaluation of prominent flare-predictive properties and methodologies, undertaken by Barnes et al. (2016), established that no single method clearly outperformed the others. This and other initial findings (e.g., the predictive value of timeseries, previous flare history) were further solidified by collaborative work on operational forecasts by Leka et al. (2019a, b) and Park et al. (2020).

Meanwhile, the explosive increase in computing power spearheaded critical advances in computer science, in an already existing Big Data ecosystem facilitated by the wealth of ground- and space-based solar observations since the mid-1990s. Data mining and the advent of machine learning eventually led to the first application of a SVM and neural networks in flare forecasting (Qahwaji & Colak, 2007). Several seminal works followed thereafter (Li et al., 2008; Qahwaji et al., 2008; Song et al., 2009; Yu et al., 2009; Bobra & Couvidat, 2015; Muranushi et al., 2015; Nishizuka et al., 2017) and the list is ever-expanding. Today, we know that solar flare forecasting – and space weather forecasting, in general – should be viewed as an interdisciplinary effort, with a potentially critical contribution from machine learning (Camporeale et al., 2018, and references therein), albeit not without open challenges impeding progress (Camporeale, 2019).

In this continuously and rapidly evolving landscape, a Consortium of nine institutes spanning over six European countries took advantage of the EU Horizon-2020 2014 PROTEC-2014 opportunity to propose the FLARECAST project. Having all the above in mind, FLARECAST pledged to *develop an advanced solar flare prediction system based on automatically extracted physical properties of solar active regions, coupled with state-of-the-art machine learning solar flare prediction methods and validated using the most appropriate forecast verification measures*. FLARECAST featured three top-level objectives, namely, one scientific, one devoted to the R2O philosophy and one engaging in communication. In particular, FLARECAST proposed:

- In terms of science, to understand the drivers of solar flare activity and improve flare forecasting.
- In terms of R2O, to provide a globally accessible flare forecast service that facilitates expansion.
- In terms of communication, to engage with space weather end users, inform stakeholders and policy makers, and educate the broader public on solar flares and space weather in general.

This collective work summarizes FLARECAST’s most significant findings and conclusions in all three objectives above, along with key elements from peer-reviewed publications that

occurred in its course. Section 2 describes the methodology followed throughout the project, while Section 3 elaborates on the tasks of data handling and monitoring. Section 4 discusses the FLARECAST performance verification strategy, while Section 5 briefly describes the project’s scientific results and explorative component. Section 6 encapsulates the main conclusions of the project’s three top-level objectives (i.e., science, operations, communication) along with lessons learned in its course. Finally, in a series of Appendices we provide detailed instructions on accessing FLARECAST data, codes and infrastructure (Appendix A), key results from the FLARECAST Users Survey (Appendix B; see Sect. 2.5.1 for a relevant discussion), the list of refereed publications related to or acknowledging the FLARECAST project (Appendix C) and, finally, a list of acronyms and abbreviations used in this paper (Appendix D).

All things considered, as commented by an attendee of one of the FLARECAST Stakeholders’ Workshops, “*the real fun starts now*”; we expect a number of future works that will take advantage of and exploit the FLARECAST products. These may not be restricted to flare prediction, as (i) the volume of metadata provided by the processing of the NRT SHARP data product (Bobra et al., 2014) is substantial and (ii) as per EU’s OpenAIRE initiative (<https://www.openaire.eu>), all data, codes and infrastructure of the project are openly available to any interested individual or team worldwide. Ideally, then, one might view the FLARECAST infrastructure as a vehicle for a future integrated space weather prediction platform. Comprehensive and diverse material on the FLARECAST project can be found in the openly accessible FLARECAST website <http://flarecast.eu> (see also Fig. 9 in Sect. 2.5 for a top-level structure of this website).

2 The FLARECAST approach

FLARECAST embraced the technical architecture and methodology structure illustrated in Figure 2. This was realized in a sequence of four procedural steps, namely, (1) [external] data acquisition, (2) property extraction, (3) machine learning-based prediction and (4) forecast verification. The overall project methodology was implemented in seven WPs, as follows:

- **WP1:** Project Management.
- **WP2:** Active Region Properties as Predictors of Flaring Activity.
- **WP3:** Flare Prediction Algorithms.
- **WP4:** Data Storage and Processing Cloud.
- **WP5:** Data and Forecast Verification.
- **WP6:** Explorative Research.
- **WP7:** Dissemination.

This work plan structure, together with the technical scheme of Figure 2, show the design philosophy of the FLARECAST science and technology. The project relied on machine learning applied to data from the HMI vector magnetograph (Scherrer et al., 2012; Schou et al., 2012) onboard the SDO mission (Pesnell et al., 2012) in order to implement a technological service for flare forecasting with the ultimate goal of contributing to a data-driven understanding of the physical triggering mechanisms of flares. Rigorous forecast verification and

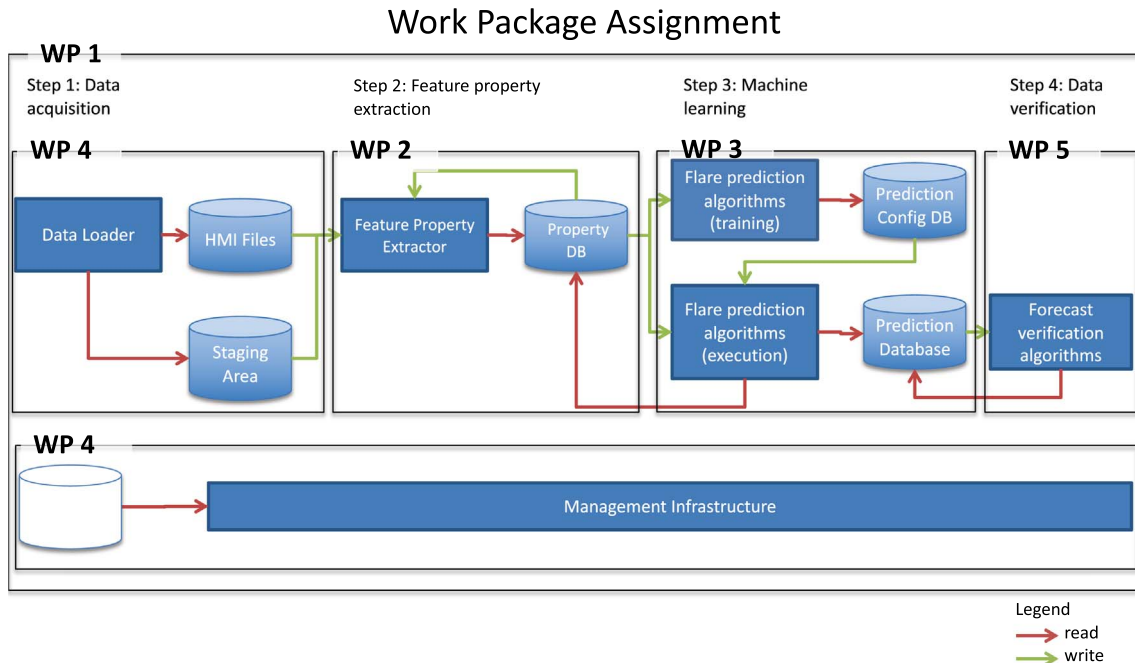


Fig. 2. The FLARECAST architecture and WP Assignment. Rectangles indicate exchangeable software components, such as algorithms and codes, while cylinders indicate data model components. WPs 6 and 7 are not included here but are integral to the project and complement the overall efforts.

dissemination of results have been crucial steps of this basic research effort.

2.1 Science motivation

The solar flare phenomenon pertains to explosive energy release in the low solar atmosphere that results in short-lasting (i.e., minutes to hours) enhancement of emission over virtually the entire electromagnetic spectrum (see, e.g., Benz, 2008; Fletcher et al., 2011 for comprehensive reviews). Major, highly-energetic flares (i.e., GOES M- and X-class events, in particular) are much less frequent (in a distribution that is long known to be a power law – see Drake, 1971; Rosner & Vaiana, 1978; Crosby et al., 1993) than minor flares (i.e., GOES C-class) and subflares⁵. In terms of mean occurrence frequency, or *climatology* in the statistical language of forecasting problems (e.g., Barnes et al., 2016; Leka et al., 2019a), major flares fall under the category of *rare* events, namely, events that are much more infrequent than the physical systems in which they appear (in this case, solar ARs). A typical 11-year sunspot cycle involves the appearance, evolution and fading of a few thousand NOAA-numbered ARs, yet Carrington’s flare is considered a one-in-a-century (or even more rare) event. In other words, viewing a set of flaring ARs as a “positive” sample vs. a “negative” sample of non-flaring ones, the ratio of sample sizes is substantially different than one. Increasing the flare magnitude threshold between positive and negative samples only pushes this ratio to further extremes. For example, in Solar Cycle 23 only ~1.8% of ARs hosted at least one GOES X-class flare (an imbalance ratio of ~0.0183), with a respective ratio of

~0.005 for GOES \geq X5.0 flares. The climatology of the FLARECAST flare sample in the even weaker Solar Cycle 24 is 1 GOES C-class flare every ~11 h, 1 M-class flare every ~4.5 days and 1 X-class flare every ~67 days; substantial variations obviously exist over different phases of the cycle. Class imbalance in major flare prediction and other rare-event problems is a central concern (Woodcock, 1976; Bloomfield et al., 2012; Jolliffe & Stephenson, 2012; Bobra & Couvidat, 2015) for machine learning methods with proposed remedies including undersampling, oversampling and misclassification weighting (e.g., Longadge & Dongre, 2013; Ahmadzadeh et al., 2021).

Flares at the top end of the observed energy distribution are viewed and treated as *extreme* events, given their very significant disruptive ability on top of their scarcity. For diverse accounts of extreme events in physical systems one may review Albeverio et al. (2006), Meyers (2011) and Sharma et al. (2012), among other comprehensive works. These accounts also refer extensively to two intrinsic characteristics of extreme events: environmental complexity and difficulty in forecasting. It is a fact that forecasting solar flares is a pressing issue for space-faring nations, mainly for two reasons: first, because of flares’ biological and technological repercussions and, second, because flares are a common element of the two other aspects of extreme space weather, CMEs and SEPs. Notwithstanding the lack of an early warning for flares, mentioned already, just a cursory examination of some biological implications indicates that, say, a 500 keV γ -ray photon has a wavelength of ~0.25 nm. This is much shorter than a DNA helix (~3.5 nm), meaning that such radiation acting on exposed humans engaging in extravehicular activities can lead to acute radiation sickness, if not being altogether fatal (see, for example, Freese et al., 2016).

The complexity, in terms of the multiscale (i.e., multifractal) behavior in the turbulent solar atmosphere (see, for example,

⁵ For a definition of NOAA/GOES flare classes, see <https://www.swpc.noaa.gov/phenomena/solar-flares-radio-blackouts>.

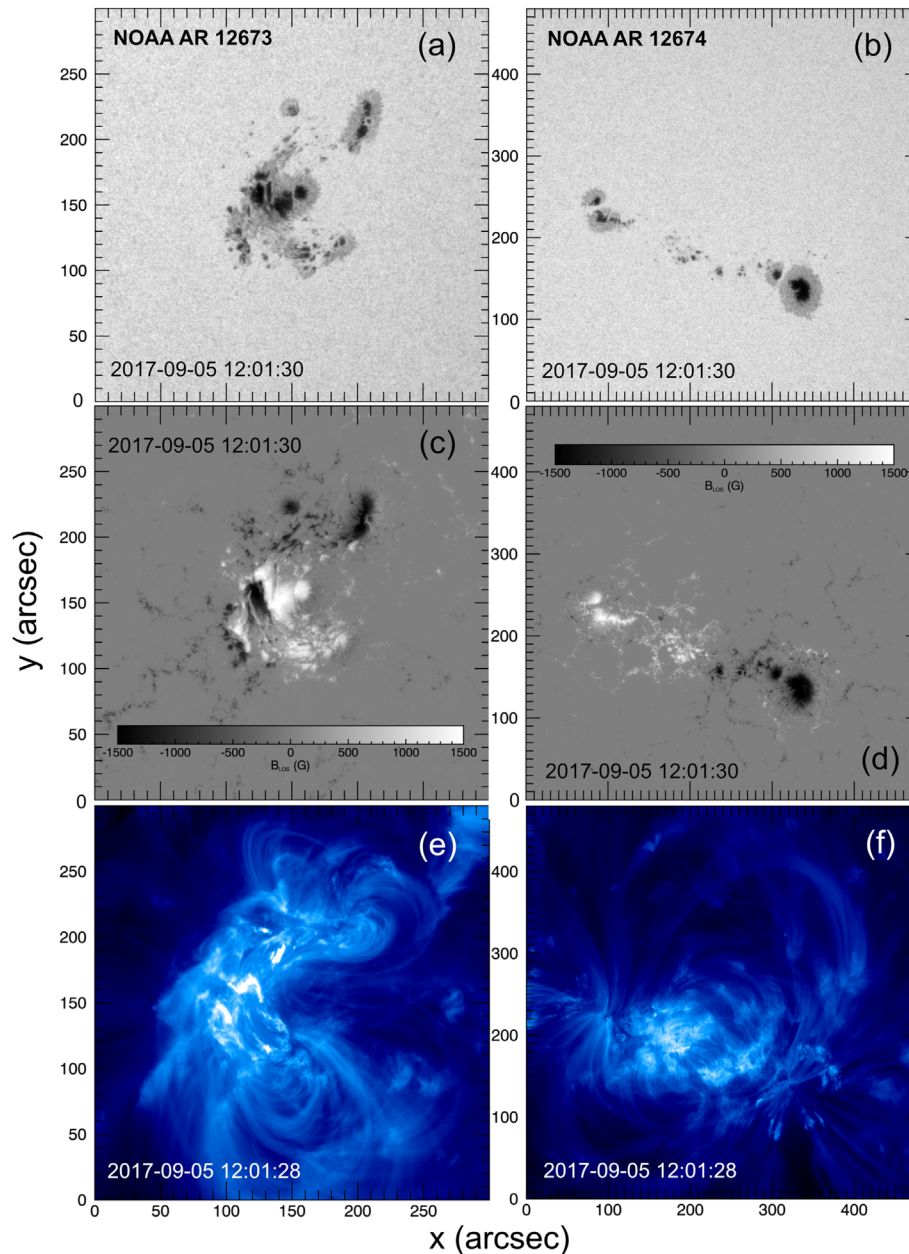


Fig. 3. Examples of two solar ARs, in terms of their photospheric continuum intensity images (a, b), LOS magnetograms (c, d) and EUV coronal images at 19.3 nm (e, f). Both active regions were observed on 5 September 2017 at around 12:01 UT. Of them, NOAA AR 12673 (left column) was intensely eruptive, giving the largest eruptive flares since January 2005, while NOAA AR 12674 (right column) did not host any major flares. Images (a–d) have been acquired by HMI, while images (e, f) by AIA, both onboard the SDO mission.

Georgoulis, 2005) undoubtedly adds to the difficulty of predicting solar flares. In particular, flares are long thought to be responses of nonlinear dynamical systems (solar ARs) in a SOC state. The SOC concept was initiated by seminal works on the topic (Lu & Hamilton, 1991; Lu et al., 1993; Vlahos et al., 1995) inspired by groundbreaking developments in theoretical physics (Bak et al., 1987, 1988; Kadanoff et al., 1989; see also Bak, 1996 for an encompassing account). An account of apparent SOC manifestations in Astrophysics, including flares, is presented in relatively recent anniversary works (Aschwanden et al., 2016; McAteer et al., 2016). However, a SOC evolution of flaring ARs would lead to an intrinsic

stochasticity in flare occurrence; this translates to an intrinsically probabilistic forecasting relying by definition on small probabilities for major and extreme flares. Randomly driven SOC models explore precisely this stochasticity; hence they are incapable of predicting small- and mid-size flares. However, Strugarek & Charbonneau (2014) reported that a class of deterministically driven SOC models could be used for predictive purposes as they raise the memory of the system, thus exerting a partial control over flare waiting times; however, this is an avenue of research yet to be sufficiently explored. Recent studies have also aimed to assess the frequency of a Carrington-level event in the framework of extreme value theory (Elvidge & Angling, 2018).

Rather than focusing exclusively on extreme events, the key scientific question and incentive behind the FLARECAST project was to determine to what extent can the skill currently achieved in the forecasting of solar flares be advanced.

Correlating solar flares with magnetic evolution dates back several decades. One of the earliest, pioneering accounts was that of Howard & Severny (1963), who reported major magnetic field changes before and after a major (coined as “great”) flare. More systematic works were added in the 1980s with limited-resolution magnetograms (Krall et al., 1982; Hagiard et al., 1984; Zirin & Liggett, 1987) linking flares –and even repeated flaring– to δ -sunspot complexes and velocity shear, before the first semi-operational flare prediction schemes appeared (McIntosh, 1990; Zirin & Marquette, 1991). In more recent years, however, the formation of long and intense magnetic PILs in the photospheric magnetic field of ARs was established as a feature of direct relevance to major flaring (for comprehensive reviews, see Schrijver, 2009; Toriumi & Wang, 2019). By “intense”, we mean PILs exhibiting substantial amounts of magnetic flux and strong magnetic gradients due to the closely seated, opposite magnetic polarities and magnetic/velocity shear (e.g., Georgoulis et al., 2019; Patsourakos et al., 2020, and references therein). Major flares occur as such PILs evolve and intensify, fueled by magnetic flux emergence and cancellation along them (van Ballegoijen & Martens, 1989; Gibson & Fan, 2006; Archontis & Török, 2008). Sometimes, an eruption including a major flare may occur in the absence of intense PILs, when the emerging flux reconnects with the pre-existing field or has enough magnetic twist to become unstable during emergence. Such an exceptional case, giving rise to a GOES X3 flare, was reported by Gary & Moore (2004). Regardless, the majority of X-class flares –the ones mostly affecting space weather conditions– occur above evolving, intense photospheric PILs.

An example of ARs with and without strong PILs is given in Figure 3, where the relatively flare-quiet (up to mid-C-class flares) NOAA AR 12674 is compared against the intensely flaring NOAA AR 12673. The latter in September 6 and 10 2017 gave the strongest (\sim X10) flares of solar cycle 24 (e.g., Yan et al., 2018). Both active regions evolved simultaneously in the solar disk, located just a few hundred arcsec away from each other. While the photospheric compactness and conspicuous PILs are evident in NOAA AR 12673, NOAA AR 12674 is much more scattered. Coronal information in Figure 3 (sampled indicatively at 19.3 nm to showcase some structure) indeed shows significant complexity in terms of several bright kernels, complexity and apparent twist; significant non-potentiality, in brief (e.g., Schrijver et al., 2005). Emission generally lacks bright kernels and does not show such non-potentiality in NOAA AR 12674.

Figure 4 further shows a PIL analysis on NOAA AR 12253, achieved in FLARECAST’s framework. More information is given in Section 5.1.

2.2 Data

2.2.1 Active region properties and flare predictors

Flare forecasting relies almost entirely on statistical correlations between solar magnetic field data and observed flare characteristics. In this case, local (i.e., AR scale) photospheric

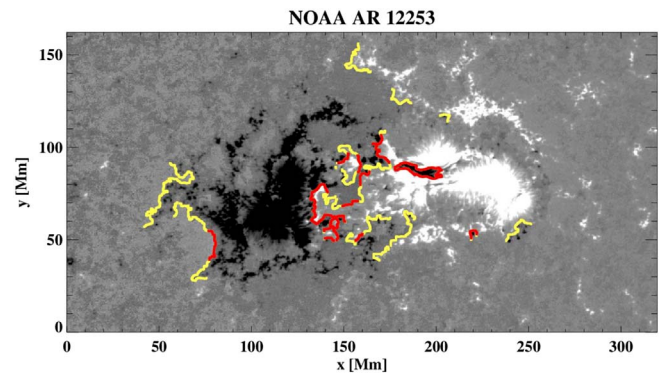


Fig. 4. Automated PIL identification in a photospheric longitudinal magnetogram of NOAA 12253, observed by SDO/HMI on 2 January 2015. The identified PILs in the region are further characterized as “moderate” (yellow contours) and “strong” (red contours). The difference between “strong” and “moderate” PILs relies on the absolute value of the gradient of the vertical magnetic field (>40 G/pixel and >16 G/pixel, respectively) and the strength of the horizontal field (>120 G and >100 G, respectively). Parts of the PIL that are not highlighted in color are below one or both thresholds and are considered as “weak”. The algorithm used is the FLARECAST PIL identification code, accessible as described in Appendix A.2.

magnetic fields are parameterized in order to identify and quantify patterns potentially associated with flares. Should a flare occur, flare detection and classification is primarily done using the GOES 0.1–0.8 nm soft X-ray band. The peak photon flux in this band is historically used to determine the GOES flare class. There are obviously other sophisticated ways to detect flares in EUV, X-ray, and optical wavelengths (see Martens et al., 2012, for an attempt in the framework of the HEK project), but the GOES X-ray classification is the one used most widely by the space weather and flare forecasting communities. Although the GOES soft X-ray detectors are, in essence, spectral irradiance instruments without spatial resolution on the solar disk, observations by the GOES/SXI telescopes complement flare information with the source ARs, at least for the largest events. However, these spatial identifications are not error-free and verification of all GOES flare locations is a nontrivial task. Significant effort on identified flare locations was also put by Hock (2012) and, more recently, by Angryk et al. (2020).

Quantitative active region photospheric properties, used as flare predictors in FLARECAST’s framework, were derived from the HMI SHARP data product (Bobra et al., 2014; Hoeksema et al., 2014). A FLARECAST processing pipeline was developed using the `hmi.sharp_cea_720s_nrt` data series that contains AR maps of the LOS magnetic field component, B_{LOS} , continuum intensity, I_c , and vector magnetic field components in the radial, B_r , co-latitude, B_θ , and azimuthal, B_ϕ , directions of the solar spherical coordinate system. These maps are produced at HMI’s JSOC⁶ at a cadence of 12 min, while the NRT HMI stream assures that these data products are available within typically 4 h of acquisition⁷ (Hoeksema et al., 2014).

⁶ http://jsoc.stanford.edu/ajax/lookdata.html?ds=hmi.sharp_720s_nrt

⁷ This time has been varying for different intervals of the SDO mission. With the current NRT data level, data are made available typically within 3 h.

Table 1. The complete list of FLARECAST properties, separated in property groups and accounting for a total of 209 predictors. The vast majority (197) of them correspond to the AR magnetic field, while the rest include information from the NOAA SWPC Catalogues and from photospheric intensity images.

Data Source	Property group	No. of predictors	Relevant predictor	Adapted from	Related references	
SWPC	Solar region summary (SRS) properties	2	McIntosh and Hale classes	McCloskey et al. (2016)	McIntosh (1990)	
		3	Number, area and longitudinal extend of sunspots		Lee et al. (2012)	
Catalogues	GOES soft X-ray flare events ^{a,b}	4	Flare magnitude, start, peak, and end times			
Surface-normal component (radial and line-of-sight) magnetograms	Effective connected magnetic field strength ^c	1	B_{eff}	Georgoulis & Rust (2007), Georgoulis (2011, 2013)		
	Fractal and multifractal parameters	1	Fractal dimension	Conlon et al. (2008)	Abramenko et al. (2003)	
		1	Generalized correlation dimension		Abramenko (2005)	
		2	Holder exponent; Hausdorff dimension		Al-Ghraibah et al. (2015)	
	Fourier and Wavelet power spectral indices	Decay index (DI) ^d	2	Power-law exponent	Hewett et al. (2008), Guerra et al. (2015)	
			8	Mean DI over PIL segments; height of DI; ratio of PIL length to DI height	Liu (2008) Zuccarello et al. (2014)	
	Magnetic PIL properties		5	Sum of PIL segments, longest PIL segment	Mason & Hoeksema (2010)	
			1	R value	Schrijver (2007)	
			1	WL_{sg}	Falconer et al. (2012)	
	3D magnetic null points ^d		6	Number of null points in different height ranges (from 2 to 100 Mm above photosphere)	Haynes & Parnell (2007)	Pontin et al. (2013)
			6	Original and partitioned Ising energy	Ahmed et al. (2010)	Barnes & Leka (2006) Kontogiannis et al. (2018)
	Full-vector magnetograms	Magnetic energy and helicity	11	Poynting flux and magnetic helicity flux proxies	Park et al. (2010) Park et al. (2012)	
		SHARP properties ^e	100	Horizontal gradient of B components; shear angle; unsigned vertical current; higher-order moments of time series	Bobra et al. (2014) (validated) Leka & Barnes (2003b, 2007)	
Magnetic energy and helicity		22	Poynting flux and magnetic helicity flux	Kusano et al. (2002)	Berger & Field (1984), Welsch et al. (2009)	
Non-neutralized currents		6	Total non-neutralized current	Georgoulis et al. (2012)	Kontogiannis et al. (2017)	
Flows around PIL		22	Speed of diverging/converging/shear flows	Park et al. (2018)	Deng et al. (2006), Wang et al. (2014)	
Intensity images ^f	Magnetic field gradient	3	Total horizontal magnetic gradient	Korsós et al. (2014)	Kontogiannis et al. (2018)	

^a AR coronal information.^b The project uses the flare attributes as properties. At least one flare of desired magnitude in the forecast window signifies a positive instance.^c Photospheric proxy for coronal information.^d Uses PFEs.^e Most SHARP parameters correspond to mean values. In the FLARECAST pipeline, other relevant magnetogram-related parameters are also used as predictors; these are maximum and minimum values, median, standard deviation, kurtosis, and skewness.^f Used in conjunction with magnetograms to calculate the sum of the magnetic field gradients between all possible opposite-polarity umbrae pairs.

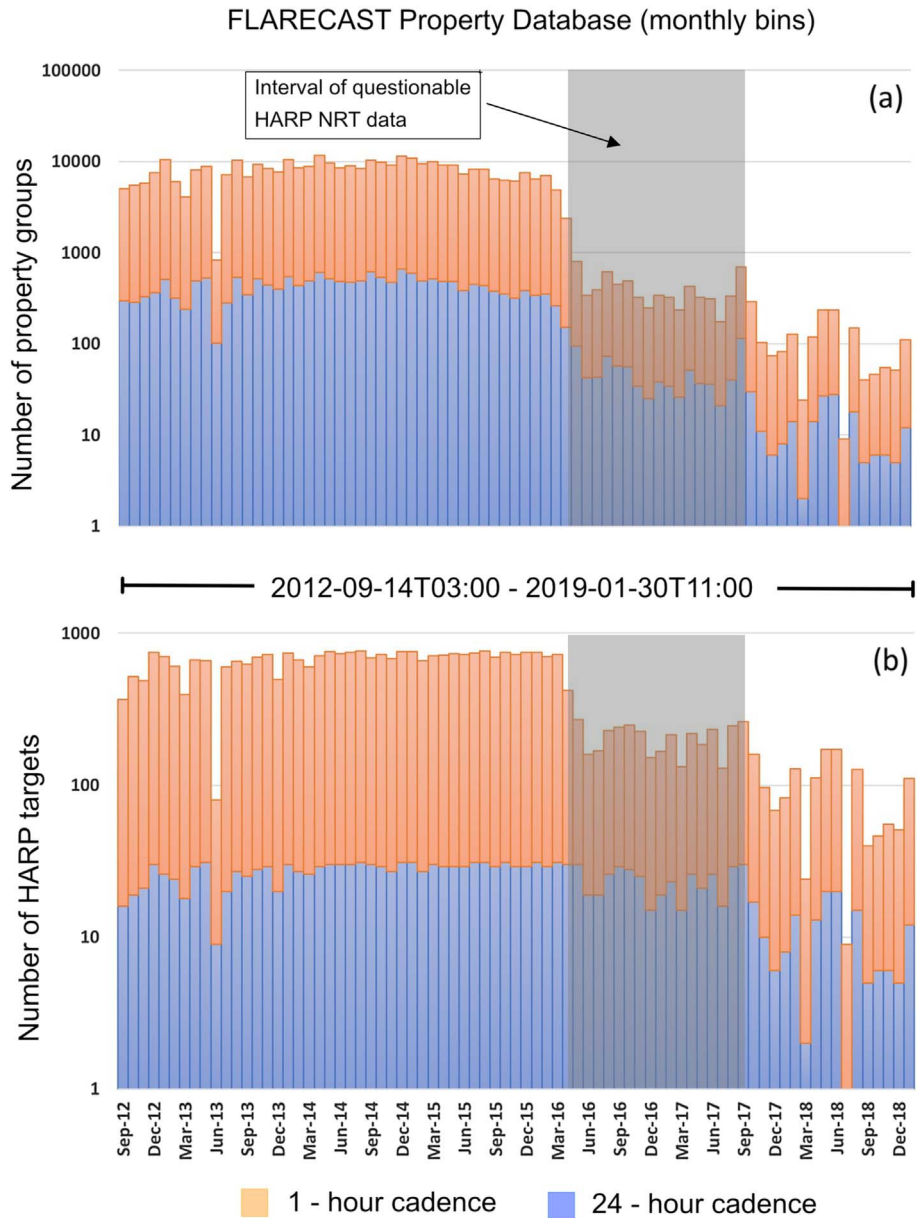


Fig. 5. Monthly numbers of (a) calculated property groups and (b) processed NRT SHARP timestamps between 14 September 2012 and 30 January 2019. Property group numbers are shown at highest (1-h; orange bars) and at lowest (24-h; blue bars) cadence. The shaded interval between 13 April 2016 and 8 September 2017 corresponds to a period of questionable quality for the HMI NRT SHARP data (see text for details). Histograms represent a total of 32,098 processed SHARP timestamps, leading to millions of predictor values, since each timestamp could potentially provide up to 209 predictors.

Another reason for using the NRT stream was that any operationally-oriented flare forecasting service has to rely on NRT data for training, testing, and validation of its method(s). In a further attempt to replicate operational conditions to the greatest extent possible, FLARECAST did not restrict its input data to longitudinal areas close to the central solar meridian, but identically treated all data regardless of solar disk position (for regions very close to the solar limbs though, some selection criteria were imposed, as discussed below). This was decided in spite of a complete understanding that magnetic projection effects, foreshortening and noise near the solar limbs are substantial.

In the FLARECAST processing pipeline, NRT SHARP maps are pre-processed before extracting any property. Pre-processing aims to:

- (a) check for missing information in the FITS headers and restore it, if possible;
- (b) examine for bad-quality or missing data (i.e., a NaN or a constant value);
- (c) capture possible differences between WCS (Thompson, 2006) positions and header positions;
- (d) trim any part of the FOV containing off-disk pixels for ARs near the limbs.

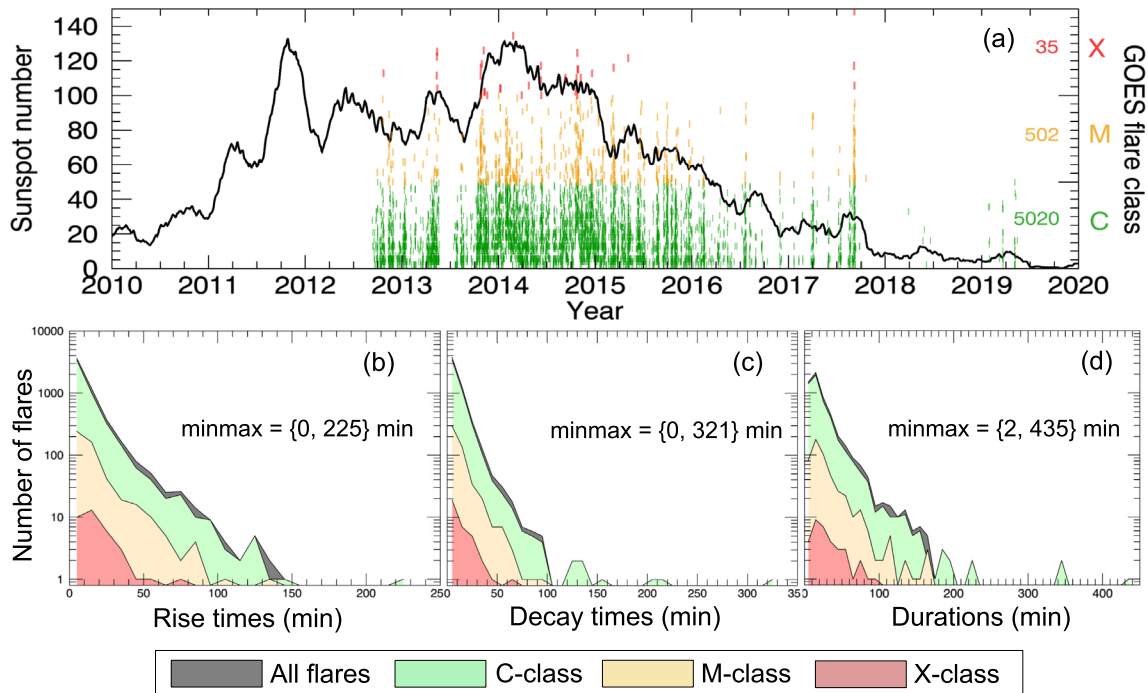


Fig. 6. Attributes of 5557 SHARP-associated flares over September 2012 to May 2019. (a) Flare onset times, in terms of GOES classes X (red), M (orange) and C (green) and respective sub-classes (tick marks; right ordinate), plotted over the 90-day averaged sunspot number over Solar Cycle 24 (left ordinate; Source: Sunspot Index and Long-Term Solar Observations [SILSO], Royal Observatory of Belgium). The numbers of flares for each GOES class are also given. The bottom line shows waterfall diagrams of the distributions of rise time (b), decay time (c) and duration (d). Extrema for each distribution are shown in each plot. The “All flares” diagram corresponds to the sum of flare numbers in each bin. We use fixed, 10-min bin sizes.

Magnetogram maps are flagged as null and no properties are calculated for SHARP timestamps and dates with (i) bad-quality data, (ii) absolute differences between FITS header coordinates and WCS-calculated coordinates that are larger than 5° or (iii) trimmed images that are less than 20 pixels in horizontal size.

Active region properties extracted from the photospheric magnetic field in FLARECAST have all been previously tested for relevance in flare prediction. These properties are presented in Table 1, itemized into 15 different groups. Each property group relates to either a certain property or a set of properties and gives rise to one or more flare predictors, respectively. These predictors, and groups thereof, quantify the state of the AR photosphere and corona and reflect (1) the entire region or SHARP FOV, (2) magnetic PILs, or (3) sunspots (as the areas with the strongest magnetic field). Most property groups characterize only the state of the photospheric magnetic field. Some, however, characterize the state of the AR corona, either by photospheric proxies or by means of current-free PFEs or by differences between the photospheric field and that expected by a PFE. The two last columns of Table 1 indicate relevant published work, distinguishing between studies that were directly implemented in the course of the project from additional or previous studies pertinent to some predictors.

It should be mentioned that, for many of the properties in Table 1, estimating uncertainties is nontrivial. For the cases uncertainties exist, mostly in terms of σ -values in fits, these values are provided in the property database. This said, prediction algorithms (Sect. 2.3) currently do not utilize predictor uncertainties. Uncertainties in verification metrics are provided,

however, as per the sensitivity analysis of Section 4.3. Evaluating the impact of predictor uncertainties in multi-parametric machine learning prediction schemes could be a meaningful future step.

Up to 209 predictors are calculated from each SHARP observation at different cadence, namely 1 h, 3 h, 6 h, 12 h, and 24 h. The full 12-min cadence of HMI SHARPs was exploited only in a very limited number of cases, due to the immense computational time it required. This computational expense dictated the different cadence as a necessity. Clearly, FLARECAST did not restrict to the 25 scalar properties included in the SHARP predictor metadata. It did, however, replicate the calculations of these predictors to validate them and consider additional image statistics (i.e., higher-order moments). Instructions on how to access these and other data and information are detailed in Appendix A.

Figure 5 displays histograms of the monthly number of property groups and SHARP timestamps analyzed over the period September 2012 – January 2019, for the finest (1 h) and the coarsest (24 h) cadences considered. Numbers corresponding to 12-min cadence are not shown due to their lack of statistical significance. The progression of Solar Cycle 24 can be roughly assessed, with significantly higher numbers until early 2016, when the declining phase gave way to the latest solar minimum. Thereafter, the numbers of both calculated property groups and processed SHARP timestamps gradually decline as eligible SHARPs in the NRT stream become fewer and fewer. The shaded interval between 13 April 2016 and 1 September 2017 corresponds to a time of potentially problematic NRT data due

Table 2. FLARECAST unsupervised and supervised machine learning methods (final release).

Unsupervised methods	<ul style="list-style-type: none"> – K-means – Fuzzy C-means – Possibilistic C-means
Supervised methods	<ul style="list-style-type: none"> – LASSO – Hybrid LASSO – Elastic net – Logit – Hybrid Logit – Random forests (RFs) – Multi-layer perceptron (MLP) – Recurrent neural network (RNN) – Support vector machine (SVM) – Garson’s method – Olden’s method

to a transient misalignment between the two HMI cameras providing filtergrams for the generation of the full Stokes vector⁸. Only definitive vector magnetogram data were reprocessed – NRTs were not – while full-disk LOS magnetograms at 45s cadence were not affected. Artifacts in this case intensify with increasing central meridian distance.

The property data set shown in Figure 5 is, to our knowledge and understanding, one of the largest and most diverse ever assembled for flare prediction, offering unique opportunities for statistical and physics-based studies to interested teams worldwide. In Guerra et al. (2018), a sub-group of six predictors were studied to determine the effects of using the LOS field (B_{LOS}) versus the surface-radial field (B_r) for their calculation and subsequent use in flare forecasting. It was determined that both LOS and radial field components can be advantageous in different circumstances. The project hence decided to use both property versions to facilitate any independent information these two versions could furnish.

2.2.2 Flare association

The NOAA/SWPC flare event list data include the universal times (UT) of their start (i.e., onset), peak, and end, along with the flare magnitude, all obtained by the GOES spacecraft 0.1–0.8 nm soft X-ray channel. Where possible, source ARs are identified from the daily NOAA/SWPC flare event lists, with their morphological properties extracted from the daily NOAA/SWPC SRS reports. NOAA ARs, flaring or not, are linked to the HMI SHARPs and the FLARECAST property database includes information on all ARs located within each SHARP FOV.

The process below is followed for each SHARP during its solar disk passage:

1. SHARPs are first checked to determine whether their FOV contains the time-advanced centroid locations of NOAA-numbered ARs. This makes use of the closest SRS report before the SHARP observation, with solar differential rotation taken into account.

2. If any NOAA ARs are assigned to the SHARP, the SWPC flare event list is searched for those NOAA numbers and related flares are associated to all of that SHARP’s property database entries.
3. For X-ray flares with no reported NOAA number, locations of co-temporal flares observed in ground-based H α images (also from the NOAA/SWPC flare event list) are used to determine if the multi-wavelength flare event occurred within the SHARP’s FOV (again taking solar differential rotation into account). Positive H α flare association results in the inclusion of the respective X-ray flare. A very small fraction of flares of GOES class C and above, of the order 0.1%, seems to miss both criteria above over the analysis period.

After all SHARP-associated flares are identified, we extract the following properties for each flare:

- F_M : GOES peak magnitude (e.g., M1.3).
- τ_s : Time difference (in seconds) between the SHARP observation time T_0 and the reported flare start time T_s (i.e., $\tau_s = T_s - T_0$).
- τ_p : Time difference (in seconds) between T_0 and the reported flare peak time T_p (i.e., $\tau_p = T_p - T_0$).
- τ_e : Time difference (in seconds) between T_0 and the reported flare end time T_e (i.e., $\tau_e = T_e - T_0$).

Figure 6 provides a snapshot of FLARECAST’s flare data set and displays the temporal distribution of different flare classes for Solar Cycle 24 (Fig. 6a), along with their rise time (Fig. 6b), decay time (Fig. 6c) and duration (Fig. 6d) for approximately the same time coverage as that of the property database statistics in Figure 5. Rise/decay times and durations have been provided by NOAA.

We consider as eligible flares only those of GOES C-class and above (i.e., $\geq C1.0$) to make sure that as many as possible are included, unobscured by an often elevated solar soft X-ray background. We acknowledge, though, that even a C1.0 flare threshold might result in loss of some flares in cases of intensely high solar activity. Others may be lost due to lack of, or an erroneous, location information.

2.3 Methods: machine learning

The FLARECAST computational component fully relied on machine learning, that is, on prediction algorithms that utilize an automatic learning step based on either labeled or unlabeled input data. The conceptual core of each machine learning technique lies on the modality of this learning step. Labeled data are characterized by a tag including one or more property-specific labels (e.g., flaring or non-flaring), whereas unlabeled data are free of such tags.

We distinguish between two broad categories of relevant machine learning methods:

- *Unsupervised methods*, that are free to infer the data structure from the data themselves and realize the learning task in a fully data-driven manner. Such methods train on unlabeled data.
- *Supervised methods*, that perform an unknown input-output mapping from known input–output samples. Sparsity enhancing techniques in these methods enable a quantitative

⁸ See https://solarnews.nso.edu/20170901/#section_hoeksema for more information.

ranking of properties (predictors) that contribute most to the achieved prediction. Supervised methods nominally train on labeled data.

The project's machine learning inventory comprises three (3) unsupervised and eleven (11) supervised machine learning methods, listed in Table 2. We chose to implement a diverse array of methods because we identified a need to experimentally test the strengths and weaknesses of each method, along with their key differences and predictive capacity. Identifying the state-of-the-art in machine learning methods flows directly from FLARECAST's main goal (Sect. 1).

2.3.1 Unsupervised methods

FLARECAST proposes three unsupervised (i.e., clustering) methods for the automatic classification of sets of unlabeled AR properties. To briefly explain how these three algorithms work, we denote as $X = \{x_k | x_k \in \mathbb{R}^d, k = 1, \dots, n\}$ a set of unlabeled samples $x_k = (x_{1k}, \dots, x_{dk})$, where x_{ik} is the i -th property of the k -th sample x_k , d is the dimension of the property space \mathbb{R}^d and $Y = \{y_j | y_j \in \mathbb{R}^d, j = 1, \dots, c\}$ is the set of the c centers of the clusters to determine. The algorithms' objective is to minimize a certain functional with respect to the cluster centers. The functionals of the K-means (Anderberg, 2014) and Fuzzy C-means (Bezdek, 1981) algorithms are given by,

$$J(X, Y, U) = \sum_{k=1}^n \sum_{j=1}^c u_{jk} \delta_{jk}^2, \quad (1)$$

and,

$$J_m(X, Y, U) = \sum_{k=1}^n \sum_{j=1}^c (u_{jk})^m \delta_{jk}^2, \quad (2)$$

respectively. In both equations above, $U = [u_{jk}]$ is the $n \times c$ membership matrix whose entries represent the memberships of the k -th sample to the j -th cluster and δ_{jk} is the distance of the k -th sample from the j -th cluster center. In the case of the K-means algorithm, the memberships are binary values, while in the case of the Fuzzy C-means algorithm they are real numbers $\in [0, 1]$, representing the membership probability. In the latter case, there is also a "fuzzifier" parameter m .

The third clustering method is Possibilistic C-means (Krishnapuram & Keller, 1996; Massone et al., 2006). This is an elaborate development of Fuzzy C-means, in which each sample can, in principle, belong simultaneously to several clusters with different degrees of membership. In this case the cost function to minimize is given by,

$$J_m(X, Y, U) = \sum_{k=1}^n \sum_{j=1}^c (u_{jk}^m \delta_{jk}^2) + \sum_{j=1}^c \eta_j \sum_{k=1}^n (1 - u_{jk})^m, \quad (3)$$

where the entries of the membership matrix satisfy the constraint $\max_j u_{jk} > 0 \forall k$ and, for each j , the regularization parameter η_j depends on the average size and shape of the j -th cluster.

2.3.2 Supervised methods

Most FLARECAST methods are supervised, in line with contemporary applications of machine learning to flare prediction (e.g., Ahmed et al., 2013; Bobra & Couvidat, 2015;

Benvenuto et al., 2018; Florios et al., 2018). A detailed description of these methods is beyond the scope of this work. The latest release of the project's platform contains a standard MLP trained by means of Error-Back-Propagation and two RNNs that allow feedback loops in the feed-forward architecture. This modification is realized by means of both an Elman neural network (Elman, 1990), in which any number of context nodes is permitted, and a Jordan neural network (Jordan, 1997) serial, in which the number of context nodes is constrained to coincide with the number of output nodes.

In a more recent development, regularization neural networks (Evgeniou et al., 2000) enable the connection between training and generalization via the minimization of functionals such as the following,

$$V(y_i, f(x_i)) + \lambda f_{\mathcal{F}} \rightarrow \text{minimum}, \quad (4)$$

where $\{(x_i, y_i)\}_{i=1}^N$ represents the training set made of N property-label pairs, $V(\cdot, \cdot)$ is the loss function that measures the price paid for the inaccuracy of predicting y_i with $f(x_i)$, and λ is the regularization parameter, which realizes the trade-off between fitting over the training set and generalization.

SVM for regression (Scholkopf & Smola, 2001) is one of the standard regularization networks implemented in FLARECAST. In this case, $V(\cdot, \cdot)$ is a standard quadratic loss function and \mathcal{F} is a Reproducing Kernel Hilbert Space (De Vito et al., 2004), in which four different kernel types can be selected (i.e., linear, polynomial, radial basis function, and sigmoidal). The FLARECAST platform also contains a SVM for classification that uses the hinge loss function, namely the one thought most appropriate for classification (Rosasco et al., 2004). Two sparsity enhancing regularization methods were also implemented, in which the number of features that effectively contribute to the generalization is constrained to the smallest possible. This is achieved by minimizing the $l1$ norm of the feature vector (i.e., the sum of the absolute values of its components), which is further achieved by means of two different approaches: penalized logistic regression (Wu et al., 2009), in which the loss function realizes the Bernoulli distribution for the labels, and LASSO (Yuan & Lin, 2006), in which the loss function is quadratic. FLARECAST also includes a hybrid version of penalized logistic regression and LASSO, in which the regression outcome is partitioned by means of a Fuzzy C-means scheme, without focusing on optimizing a specific skill score (Benvenuto et al., 2018). The platform contains a further generalization, namely an elastic net (Zou & Hastie, 2005) algorithm, in which the minimization functional contains two penalty terms ($l1$ and $l2$) with two different regularization parameters optimized by cross validation.

Ensemble learning (Dietterich, 2000) is another supervised approach that uses a combination of different learning models to increase the classification accuracy. In this framework, FLARECAST offers a RF algorithm (Breiman, 2001), which works as a large collection of de-correlated decision trees. Given a training set of samples made of properties and corresponding labels, a decision tree recursively splits the training samples into subsets based on the value of a single property. Each split corresponds to a node in the tree and the task is to separate records in the training set that have different characteristics. We follow the implementation described in Liaw & Wiener (2002), by splitting the tree until every subset is pure (i.e., all samples in the subset belong to the same class). In this way all terminal nodes (i.e., the leaves) are assigned a unique class label. Once

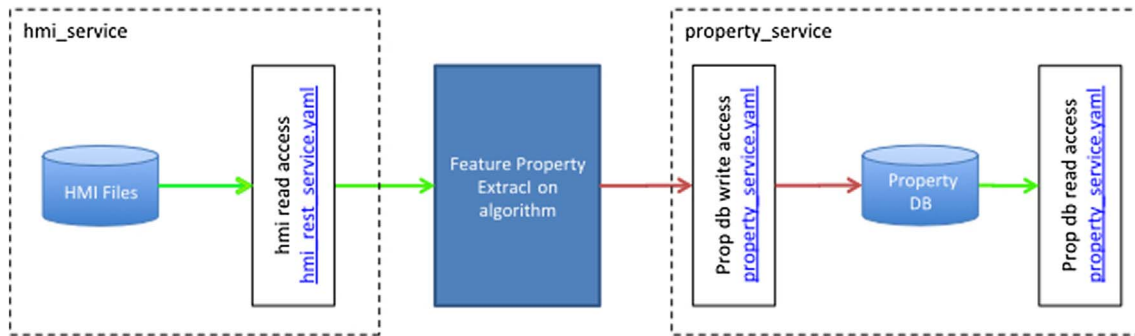


Fig. 7. A data component (dashed box) consists of a database (blue cylinder) and a well-defined API (white box). An algorithm (blue rectangle) uses this API to read data (green arrow) and another API to write data (red arrow) into a separate data component.

the decision tree has been constructed, classifying a test record is achieved by starting from the root node, applying the test condition to the record and following the appropriate branch based on the outcome of the test. This can lead either to another internal node, for which a new test condition is applied, or to a leaf node. If a leaf node is reached, the label associated with it is assigned to the record. In the RF approach, the training set is randomly divided into a fixed number of subsets and for each subset a decision tree is built. New, incoming unlabeled samples are classified by aggregating the predictions of the decision trees via a majority vote procedure.

Details on how to obtain the FLARECAST machine learning algorithms can be found in [Appendix A](#). We note in passing that the landscape of machine learning methods is continuously and rapidly evolving, with new algorithms constantly introduced. Our approach, as explained in [Section 2.4.3](#) below, is an implementation based on modularity in such a way that incoming machine learning methods can be easily integrated in the FLARECAST platform.

2.4 Technology

FLARECAST relies on a broad selection of different technologies, including hardware for storage and computing as well as software for data handling, infrastructure management and computation. The software was designed in a way that is hardware independent and can be installed on numerous platforms. All code developed during FLARECAST has been published under an open source license and is freely available. Code acquisition and license information are detailed in [Appendix A](#).

2.4.1 Computing hardware

A computing server dedicated to FLARECAST has been integrated into the MEDOC computing infrastructure⁹ and hosts the production version of the FLARECAST pipeline. Queries on this server for SDO/HMI data are made directly into the MEDOC database tables, and SDO/HMI files are accessed locally. This allows efficient runs of the FLARECAST property extraction algorithms.

2.4.2 Data storage hardware

The main FLARECAST data volume is provided by eight SDO/HMI data series at 12 min cadence that have been

downloaded using the SDO NetDRMS and archived as part of the MEDOC solar physics data archive. Besides the one used primarily for the forecasting tasks (`hmi.sharp_cea_720s_nrt`), downloaded series included `hmi.m_720s`, `hmi.m_720s_nrt`, `hmi.sharp_720s`, `hmi.sharp_720s_nrt`, `hmi.sharp_cea_720s`, `hmi.ic_720s` and `hmi.ic_nolimbdark_720s_nrt`. Not all of them were finally used by the Consortium, although definitive SHARP data and continuum images were used for tasks akin to the explorative science component ([Sect. 5](#)). NRT data series were downloaded no later than 1 hr after they were made available by HMI's JSOC. This overall delay could be further reduced, if necessary, by more frequent download requests.

2.4.3 FLARECAST software architecture

The FLARECAST architecture design was driven by the needs for modularity, portability and ability to perform and accommodate different algorithms written in various programming languages. This philosophy is best described by the top-level diagram of [Figure 2](#) in [Section 2](#). The diagram reflects the four processing steps of the FLARECAST pipeline:

- **Step 1:** Acquisition and transformation of data from multiple sources (SDO/HMI & NOAA SWPC – [Sect. 2.2](#)).
- **Step 2:** Extraction of properties from the data by several algorithms ([Sect. 2.2](#)).
- **Step 3:** Prediction through implementation of several machine learning algorithms ([Sect. 2.3](#)).
- **Step 4:** Verification of the generated forecast data products ([Sect. 4](#)).

The underlying management infrastructure controls and monitors these algorithms. The layout of the FLARECAST architecture simplified the project management as the individual components are under the responsibility of dedicated WPs. The relation between components and WPs is shown by the black rectangles in [Figure 2](#). The software infrastructure is discussed in detail in [Section 3](#).

2.4.4 Software components

The software components, represented by the blue rectangles in [Figure 2](#), fulfill specific tasks such as data loading, property extraction, machine learning, or verification. Generally,

⁹ <https://idoc.ias.u-psud.fr/MEDOC>

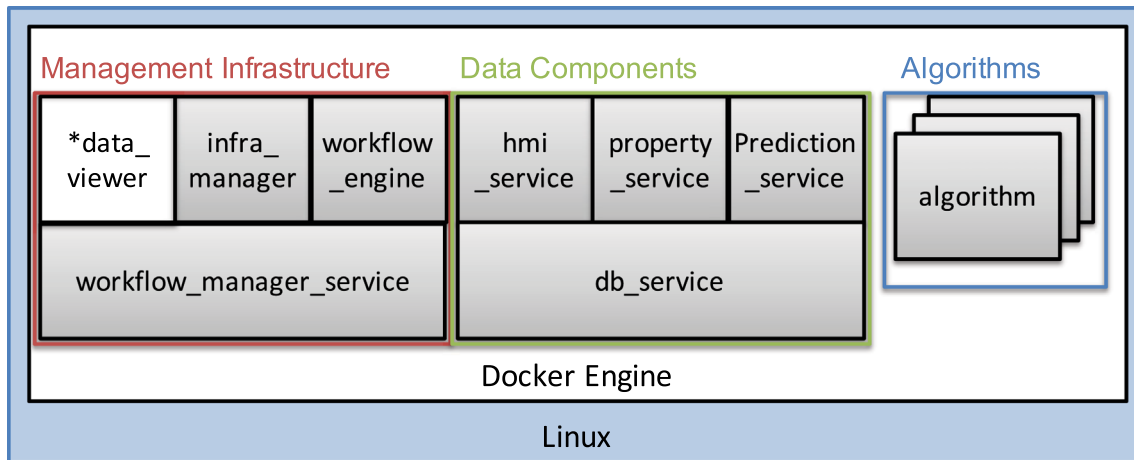


Fig. 8. Components of the FLARECAST architecture within the main Docker engine. The different grayscale rectangles correspond to different Docker containers, while adjacent containers typically share a common interface.

a software component contains several algorithms and each algorithm implements a specific variant of a task. Each property extraction or machine learning method, for example, has its own dedicated algorithm.

The management infrastructure orchestrates the execution of the FLARECAST workflow. It launches the extraction algorithms as soon as new observational data arrives. After they are finished, the management infrastructure triggers the flare prediction algorithms.

The management infrastructure is implemented as a collection of containers (see Sect. 2.4.6 below). In automated mode, a set of repeatedly executed computer scripts (i.e., cron jobs) regularly check for new data and start the algorithms, if necessary. Additionally, a small web application allows users with administrator privileges to manually trigger the execution of algorithms.

2.4.5 Data components

The blue database symbols in Figure 2 denote individual parts of the FLARECAST data model, hereafter referred to as data components. Within the workflow, every software component (i.e., every algorithm) reads from several and writes to precisely one data component. Like this, the software components are decoupled from each other and only have to provide a control interface (start, stop) for the management infrastructure. The management infrastructure connects to a dedicated data component for configuration and data logging.

Each data component defines a generic interface to read, create, update, and delete data. The read methods include a query language for simple queries. Figure 7 illustrates this process for an example property extraction algorithm.

2.4.6 Docker containers

All FLARECAST infrastructure components are implemented as Docker containers. Docker¹⁰ is an open source software ecosystem (i.e., engine) in which different Docker

containers co-exist and function independently. A Docker container can be viewed as a light-weight virtual machine that hosts an arbitrarily configured software environment. This includes custom programming languages and different library versions per container. Each FLARECAST infrastructure component or algorithm is deployed in a dedicated Docker container. These containers can be installed on a high performance cluster as well as on a developer's desktop machine, making it possible to deploy FLARECAST in different environments at the same time. A schematic of the FLARECAST Docker containers is shown in Figure 8.

2.4.7 Language independence

As mentioned already, algorithms may be written in any programming language. This allows to re-use already existing libraries for certain tasks. Currently tested and supported are Python 2.7 and 3.0, the Interactive Data Language (IDL) 8.5 and R. Other languages that can be executed on a Linux system (e.g., Java, C++ and Perl) are not expected to generate major issues, but we must make clear that we have not systematically tested them.

2.4.8 Data model

The design of the data model is balanced between flexible support for different kinds of data structures and strict schema definitions for better interoperability. This can be achieved through a semi-structured data model. Its main structures are defined by a traditional table-based, relational data model. Individual entries of the table support weakly-typed data types (i.e., types supported by weakly- or loosely-typed programming languages, such as C). In this sense, an algorithm can handle data types as simple as numeric values or strings simultaneously with more complex data types, such as arrays, matrices, dictionaries and whole hierarchies of objects. For the weakly-typed data types we rely on the built-in JSON data type of PostgreSQL 9.6. Algorithms encode their output (e.g., extracted properties) in the JSON data format. JSON is supported by most C-like programming languages, including Python, R and IDL.

¹⁰ <https://www.docker.com>

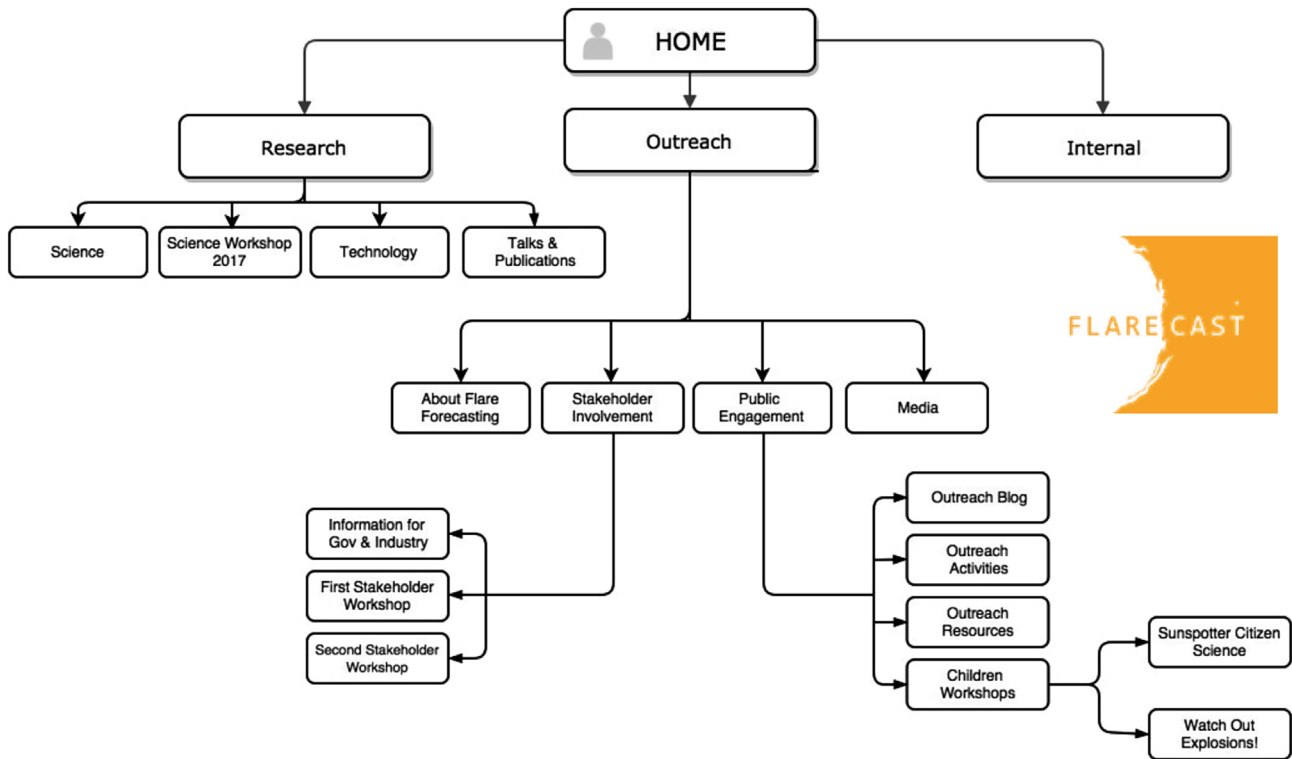


Fig. 9. Structure and components of the FLARECAST website (<http://flarecast.eu>).

2.5 Communication and dissemination

The FLARECAST communication objective ran throughout the project as it had a dedicated, overarching WP (WP7; see the WP list in Sect. 2). The FLARECAST Consortium aimed to raise awareness in three different directions, namely industry and government, space weather end users, and the general public.

2.5.1 Industry/government, scientists and end users

Initial engagement with end users, in terms of both scientists and forecasters, evolved into a User Survey conducted by the Met Office partner in autumn 2016, in advance of the FLARECAST First Stakeholders’ Workshop held at the Met Office (Exeter, UK) in January 2017. In its findings (described in more detail in Appendix B) about a quarter of respondents were not sure about the accuracy of the forecasts they were using, clearly indicating a need for more education about the forecast methods, as well as about the verification of data, methods, and performance. This may also explain why around 60% of respondents were passive regarding recommending forecast services. Further, it was found that timeliness, accuracy and ease of use are the most important factors in a forecast, while scientific detail is the least important information to include.

The First Stakeholders’ Workshop was attended by 30 participants in total, including 20 non-FLARECAST attendees from basic solar and space-weather research and operational forecasting, as well as end users and stakeholders from the aviation, defence, marine, satellite, and communications sectors. There was discussion of the survey results and an extended

discussion on the construction of rough research and development roadmaps for the short and long terms. Top-level conclusions of the Workshop included the following:

- One forecasting solution is not possible to fit all sectors and users, as different sectors need different forecast windows, latencies and means of verification depending on their generic business models (i.e., false alarms, misses, cost-to-loss ratios, etc. . .).
- FLARECAST forecasts will need an intermediate step before reaching end users, to make the forecasts understandable to them. This step should be taken by operational centres, such as the Met Office Space Weather Operations Centre (MOS-WOC). It was deduced that the real end users of FLARECAST are, in fact, the operational forecasters themselves.
- The users stressed the practical aspects of an operational forecast (i.e., reliability and precision) over scientific details. They also found the Workshop to be an excellent educational experience.
- The need for an integrated, Sun-to-Earth, space-weather forecasting system was portrayed as evident and pressing. It will become increasingly pressing in the future.

A Second Stakeholders’ Workshop was held in Ostend (Belgium) in November 2017 (alongside that year’s European Space Weather Week) and was oriented more toward scientists, end users and operational space-weather forecasters. Among others, it featured participation by the EASA and NASA’s CCMC. This second workshop pointed out that FLARECAST is producing and publishing results that are readily available to be used and validated, and the FLARECAST forecasting service

will follow shortly. However, some convenors agreed that, in some respects, the “fun” for FLARECAST’s potential really starts now, since the greater community should think about scientific/operational projects involving the new FLARECAST “software instrument”. Finally, in view of the strong contribution from EASA in the discussion, it is increasingly clear that aviation agencies (and the aviation industry, naturally) are becoming more engaged in space-weather impacts and associated forecast services. Therefore, the heliophysics community needs to proactively follow up with them to ensure that this opportunity for a meaningful, prolific interaction is not missed.

In addition, intense communication took place with the scientific community at conferences and during a dedicated FLARECAST Science Workshop organized in June 2017 in Paris by the CNRS partner. These actions anchored FLARECAST within the larger field of current and future space-weather research and secured collaboration between interested parties beyond the project’s duration.

2.5.2 Education and public outreach

A variety of communication formats were used to involve the public in all partner countries: science cafés; citizen science events; children’s workshops; interactive exhibits; social media. Public engagement activities raised awareness not only about space weather and the need for reliable solar flare forecasts but also, more generally, about the working of collaborative research in Europe. We witnessed the relentless interest and fascination of the public on the realization that real efforts are made to predict solar activity, because this can have a very tangible impact on human life and well-being. This was not initially clear and we strove to pass the word clearly and compellingly.

2.5.3 Press releases and press appearances

The numerous direct interactions and networking activities between scientists and different stakeholders in the project’s framework were aimed at sharing information, advancing mutual understanding, and increasing trust in science (i.e., embedding the project in society at large). These efforts came just in time, as two partner countries (Switzerland and United Kingdom) faced political uncertainties about their involvement in future European research programmes. In one remarkable development¹¹, the project itself was brought up as a paradigm of solid, meaningful collaboration between British and Continental European researchers.

This visibility was assisted by the Consortium’s concerted efforts to exploit opportunities for disseminating the project’s achievements and deliberations. There were annual press releases in all official languages of the partner countries, while Consortium members were interviewed in numerous occasions by the electronic and printed media based in the partner countries. Much attention was, therefore, paid to contextualized press work to make sure that the project was broadly communicated in all participating countries and beyond.

Detailed information on all three communication and dissemination elements of FLARECAST can be found on the project’s website, a skeleton diagram of which is provided in Figure 9.

¹¹ <https://www.theyworkforyou.com/lords/?id=2016-11-03b.771.2>

3 The FLARECAST infrastructure: handling and monitoring

The software developed within FLARECAST is a computational resource enabling one to apply several machine-learning methods to data available in the FLARECAST property database. The software is object-oriented, with a set of objects for reading and analyzing data formats (i.e., model learning or prediction), another set for visualizing the results and, finally, another set for storing them in appropriate prediction databases. The software has been designed and written in Python, with the pertinent software modules described in the following sections.

3.1 Data handling

Data handling is managed by a database interface module that can read and collect properties from the FLARECAST property database and write the results of the analysis into other databases. The databases involved are:

1. **Property database:** It contains the relevant AR information and comprises the following data:
 - (a) SHARP HMI metadata (i.e., properties);
 - (b) data set of SHARP-calculated properties exclusive to FLARECAST;
 - (c) NOAA/SWPC SRS data;
 - (d) flare association as extracted from the NOAA/SWPC event list.
2. **Machine learning configuration database:** It contains information learned by a given machine learning method from a given training set according to the machine learning algorithm category. We have generated the following machine learning algorithm categories and corresponding parameters to be saved and stored:
 - (a) Neural networks → architecture, synaptic weights.
 - (b) Clustering methods → clusters’ centers.
 - (c) Regression methods → predictors’ weights, regularization parameters.
3. **Machine learning result database:** It contains the prediction results associated to a given AR property set (i.e., the labels predicted by the selected machine learning algorithms). Depending on the machine-learning algorithm utilized, ARs at a given point in time can be labeled using different types of prediction labels. The label types are:
 - (a) Predicted flare occurrence: binary values, 0 (no-flare)/1 (yes-flare) values, referring to the flare occurrence at a specific GOES flux level that must be fixed a priori (e.g., flares \geq M1.0).
 - (b) Predicted flaring probability: decimal percentage values between 0 and 1, similar to above.
 - (c) Flare intensity: a positive real number.
 - (c) Flare intensity and delay: two positive real numbers.

The flare intensity number reflects the GOES flare class (forecast and actual), while the delay refers to the interval between the time of the specific flare forecast and the actual flare onset.

3.2 Data monitoring

A consistent and as complete as possible coverage of the SDO/HMI archive at MEDOC is important for FLARECAST to fully exploit the JSOC data, along with the ability to promptly download the latest SDO/HMI NRT SHARPs for as long as they are available. As a result, both the archive coverage and download delays had to be monitored. Additional monitoring tools have, therefore, been developed to keep track of both the infrastructure workflow and database coverage.

3.2.1 Data coverage monitoring

Monitoring of data coverage is achieved using the scripts in the “coverage” directory of the project’s public Git repository.¹² These scripts are automatically executed daily with their results published at MEDOC.¹³ This webpage uses color tables to display the percentage of existing SDO/HMI data, for different data series, that are available at JSOC and are mirrored at MEDOC. This is done as a function of month and day of month using a color table in which even just one missing observation is visible and encodes in white those cases where no data are available at JSOC. Notice that for the SHARP data series a gap may exist either when no eligible SHARPs can be found on the solar disk on a given day or in (relatively rare) cases of issues with the HMI data pipeline.

3.2.2 Download delays monitoring

Monitoring of download delays is achieved using the script in the “delays” directory of the above-mentioned Git repository. The observation date and time at which files are available at MEDOC are directly queried from the database, with year and month supplied to the script as external arguments.

The output is a plot displaying the download delay (compared to the observation [i.e., data acquisition] time) as a function of observing time for the selected month. The delay axis spans from 0 h to 24 h. Missing data (that could not be downloaded from JSOC) are shown as a 0-h delay in red, while data downloaded with a delay of more than 24 h are shown as a 24-h delay in orange.

3.3 Infrastructure and database coverage monitoring

To overview and monitor the workflow process and database coverage, the FLARECAST infrastructure provides several viewers, some of which are intended as internal tools. These tools can be found in dedicated Git repositories in the INFRA project area.¹⁴ The *infra_viewer* allows FLARECAST developers to manually start new processes and to monitor their performance. The *property_viewer* and the *prediction_viewer* allow to visually inspect the distribution and completeness of the data. They help system admins to better assess the quality of the data.

In addition to viewers, the FLARECAST infrastructure provides a collection of API routes to analyse the data on a database level (Appendix A). These scripts iterate through the entire

Table 3. 2×2 contingency table for categorical flare forecasting.

		Flaring observed	
		Yes	No
Flaring Predicted	Yes	True positive (TP)	False positive (FP)
	No	False negative (FN)	True negative (TN)

data set and report unexpected gaps in text form. Manual analysis is then needed to decide if data need to be reprocessed or the gaps can be explained by missing or corrupted source data (i.e., HMI SHARPs).

4 Performance verification strategy

The entire verification software for the FLARECAST project is contained in the *verification_engine.py* and *verification_module.py* codes in the project’s Stash repository.¹⁵ The starting point for the software is reading from a JSON-format configuration file. This initializes the algorithm name for the prediction method, the threshold flare intensity of interest, the base issuing time of the forecasts, the length of the forecast window, the latency of the forecast windows from their issuing times, whether magnetic and/or flare history property sets are used and, finally, the start and end date-times for the period undergoing verification. Given an input forecast window length, the software determines which issuing times will be considered for reading prediction “algorithm_config_name” entries from the prediction database. This is to ensure that temporally edge-to-edge forecasts are considered in the verification, such as a single 24-h forecast window issued at 00:00 UT each day, or two 12-h forecast windows issued at 00:00 and 12:00 UT each day, or four 6-h forecast windows issued at 00:00, 06:00, 12:00, and 18:00 UT each day. In all of these examples, the latency (i.e., the time difference between a forecast issuing time and the start of the prediction window) is zero, meaning that forecasts are effective immediately. Nonzero latency can also be implemented, if desired.

Following this, two separate calls are made to the FLARECAST prediction API. The first is a request to the “algoconfig/data” service for the related prediction algorithm configuration settings (or a list of prediction algorithm configuration settings in the case of forecast window durations less than 24 h). The second is a request for the entire set of prediction results through the “prediction/list_v2” service. Having retrieved the full set of forecasts to be verified, minimum and maximum flare class information for that prediction set are extracted from the “algoconfig/data” results. This information is needed to convert the associated flare information contained in each prediction result entry into the relevant binary observational truths (i.e., one or more suitable flare did [1], or did not [0], happen within the forecast window) that are then compared to the decimal forecast probability or binary classification issued by the prediction algorithm under consideration.

The next stage of *verification_engine.py* is the initial construction of the output metadata structure, containing the full set of metadata keys and a placeholder for the “data” key. At this

¹² <https://dev.flarecast.eu/stash/projects/INFRA/repos/hmi-coverage/>

¹³ <http://sdo.ias.u-psud.fr/medoc-hmi-list/>

¹⁴ <https://dev.flarecast.eu/stash/projects/INFRA>

¹⁵ https://dev.flarecast.eu/stash/projects/VER/repos/standard_verification

Table 4. Metrics and skill scores used for categorical forecasts in FLARECAST, along with their applicable ranges. All parameters shown rely on the simple 2×2 contingency table of Table 3.

Name	Notation	Formula	Range
Accuracy	ACC	$\frac{TP + TN}{N}$	[0, 1]
False alarm ratio	FAR	$\frac{FP}{TP + FP}$	[0, 1]
Bias	BIAS	$\frac{TP + FP}{TP + FN}$	[0, ∞]
Threat score	TS	$\frac{TP}{TP + FN + FP}$	[0, 1]
Equitable threat score	ETS	$\frac{TP - R_{ETS}}{TP + FN + FP - R_{ETS}}$ Using $R_{ETS} = \frac{(TP + FN)(TP + FP)}{N}$	$[-\frac{1}{3}, 1]$
Probability of detection	POD	$\frac{TP}{TP + FN}$	[0, 1]
Probability of false detection	POFD	$\frac{FP}{FP + TN}$	[0, 1]
Odds ratio	OR	$\frac{TP \cdot TN}{FN \cdot FP}$	[0, ∞]
Odds ratio skill score	ORSS	$\frac{(TP \cdot TN) - (FN \cdot FP)}{(TP \cdot TN) + (FN \cdot FP)}$	[-1, 1]
Heidke skill score	HSS	$\frac{TP + TN - R_{HSS}}{N - R_{HSS}}$ Using $R_{HSS} = \frac{(TP + FN)(TP + FP) + (TN + FN)(TN + FP)}{N}$	[-1, 1]
True skill statistic	TSS	POD - POFD	[-1, 1]
Symmetric extremal dependence index	SEDI	$\frac{\log(\text{POFD}) - \log(\text{POD}) - \log(1 - \text{POFD}) + \log(1 - \text{POD})}{\log(\text{POFD}) + \log(\text{POD}) + \log(1 - \text{POFD}) + \log(1 - \text{POD})}$	[-1, 1]
Appleman's discriminant	AD	$\frac{TN - FN}{FP + TN}$ if $(TP + FN) > (FP + TN)$ $\frac{TP - FP}{FN + TP}$ if $(TP + FN) < (FP + TN)$	$[-\frac{FN}{FP}, 1]$ $[-\frac{FP}{FN}, 1]$

point, “forecast_type” and “source_predictions” are left blank as the content depends on whether forecasts are probabilities or classifications; distinction is not possible based only on “algorithm_config_name” because some prediction methods generate both types. Classification forecast-observation pairs are then passed into the *calc_class_stats()* function in *verification_module.py*, while probability forecast-observation pairs are passed into the equivalent *calc_prob_stats()* function. Both return structured results saved directly into the “data” key at the metadata level of the output structure. Numerous skill

scores are then computed for both the probabilistic and the categorical-classification format of forecasts, outlined in more detail below.

4.1 Probabilistic forecasting

The computation of probabilistic scores relies on two metrics associated to the i -th forecast: the binary observed flaring $o_i \in \{0, 1\}$, and the forecast probability $f_i \in [0, 1]$. Using them, the verification module computes the Brier Score,

$$BS = \frac{1}{N} \sum_{i=1}^N (o_i - f_i)^2, \quad (5)$$

together with its decomposition into the three components of reliability, resolution, and uncertainty (see, e.g., Richardson, 2012). From the BS, one obtains the Brier Skill Score,

$$BSS = 1 - \frac{BS}{BS_{\text{ref}}}, \quad (6)$$

where the reference score BS_{ref} is given by,

$$BS_{\text{ref}} = \frac{1}{N} \sum_{i=1}^N (o_i - \bar{o})^2, \quad (7)$$

and $\bar{o} = \frac{1}{N} \sum_{i=1}^N o_i$ is the average of the binary observed flare occurrences over the N forecasts (i.e., the climatology).

Notice that BSS directly combines probabilistic forecasts with binary observations in a paired format. Further characterization is achieved through binning of these forecast-observation paired data according to their forecast values. This enables construction of the four quantities required to plot reliability diagrams, which are plots comparing the observed frequency of events to the forecast probability (see, e.g., Broecker, 2012, as well as Murray et al., 2017 for an application to flare forecasting). These quantities are (1) the number of entries in each forecast bin, (2) the average forecast value in each bin, (3) the average observation in each bin, and (4) an estimate of uncertainty in the average observation value. The key parameter in the construction of these reliability diagram quantities is the size of the decimal-probability forecast value bins, which is controlled by the “data set”–“probabilities”–“rel_dia_stepsize” parameter in the JSON configuration file input into *verification_engine.py*. The reliability diagram data are specifically included in the verification output “data” result structure to prevent the need for the full prediction set to be requested again through the API service at a later time.

In addition to the standard probabilistic metrics, decimal forecast probabilities are converted by a threshold value into categorical-classification forecasts of *no-flare* (i.e., probabilities below the threshold) and *yes-flare* (i.e., probabilities at or above the threshold). Varying this threshold yields a series of binary forecast and observation data sets that have the full set of categorical metrics and skill scores calculated for them. These threshold-dependent categorical metric data sets are then included in the output probability verification “data” result structure as a nested list of categorical-classification output result dictionaries, with each list entry supplemented by the threshold probability value used to construct that entry. These categorical metrics and skill scores are provided below.

4.2 Categorical forecasting

Given a certain range of flare magnitudes, forecast window duration and latency, verification of the simplest possible categorical forecasting scheme (i.e., binary 0 for *no-flare* and 1 for *yes-flare*) relies on the 2×2 contingency table (aka “confusion matrix”) of Table 3. This is constructed from the intersection of the binary forecast-observation pairs corresponding to the following four cases: number of forecast windows predicted to flare with flare(s) observed (TP); number of forecast windows

predicted to flare with no flare(s) observed (FP); number of forecast windows predicted to not flare with flare(s) observed (FN); number of forecast windows predicted to not flare with no flare(s) observed (TN). The first row (TP + FP) provides the number of flare-predicted forecast windows, while the second row (FN + TN) corresponds to the number of non-flare-predicted forecast windows. The first column (TP + FN) provides the number of observed flaring forecast windows, while the second column (FP + TN) corresponds to the observed non-flaring forecast windows. The total number of forecast windows is $N = TP + FP + FN + TN$. Table 3 gives rise to a long list of possible forecast metrics and skill scores, with those utilized in FLARECAST outlined in Table 4.

Another important element of forecast verification is the use of the POD and POFD (see Table 4) as ordinate and abscissa, respectively, of the ROC plot (see, e.g., Broecker, 2012, as well as Sharpe & Murray, 2017 for an application specific to flare forecasting). For thresholded probabilistic forecasts (Bloomfield et al., 2012) the ROC plot contains the (POD, POFD) pairs for each probability threshold. In a manner similar to the reliability diagram of Section 4.1, readily creating ROC plots avoids unnecessary reprocessing of large quantities of forecast data just for visualization purposes. ROC plots directly provide the AUC metric that is determined by the integration of the POD values over their spacing in POFD.

4.3 Verification metric uncertainty calculation

In standard operation, the verification process calculates a single value of each metric and skill score that represents the performance achieved across all forecasts within the selected verification time range. FLARECAST also includes an optional assessment of uncertainty for these single-valued metrics and skill scores, controlled through the “data set” – “uncertainties” parameters in the JSON verification configuration file. The choice of resampling method for the calculation of uncertainties is controlled by the string-format “sampling_scheme” parameter. In the current implementation, three resampling schemes can be utilized, namely bootstrap and jackknife (Efron, 1982) or random sampling without replacement (i.e., a sub-sampling method). The size of each partial sample is set via “sample_percent”, namely the decimal percentage of the total number of forecast-observation pairs in the verification time range, N . Finally, the number of times that resampling is performed is set by the “realizations” parameter. It is worth noting that the jackknife method overrides these to self-define the sample size as $N - 1$ (one less than the number of forecast-observation pairs) and the number of realizations as N , because this is the number of times one forecast-observation pair can be dropped. In addition, the standard bootstrap method uses N samples (i.e., “sample_percent” = 1), with the freedom to choose any number of realizations. In our infrastructure it is also possible to run the FLARECAST verification engine with the bootstrap resampling with replacement but using “sample_percent” < 1, although this should not be treated (or referred to) as a standard bootstrap implementation, except in the case “sample_percent” = 1.

The assessment of metric uncertainties is optionally invoked after the categorical-classification or probability forecast metrics have been written into the output verification data structure. In the *calc_ss_uncert()* uncertainty assessment function, indices

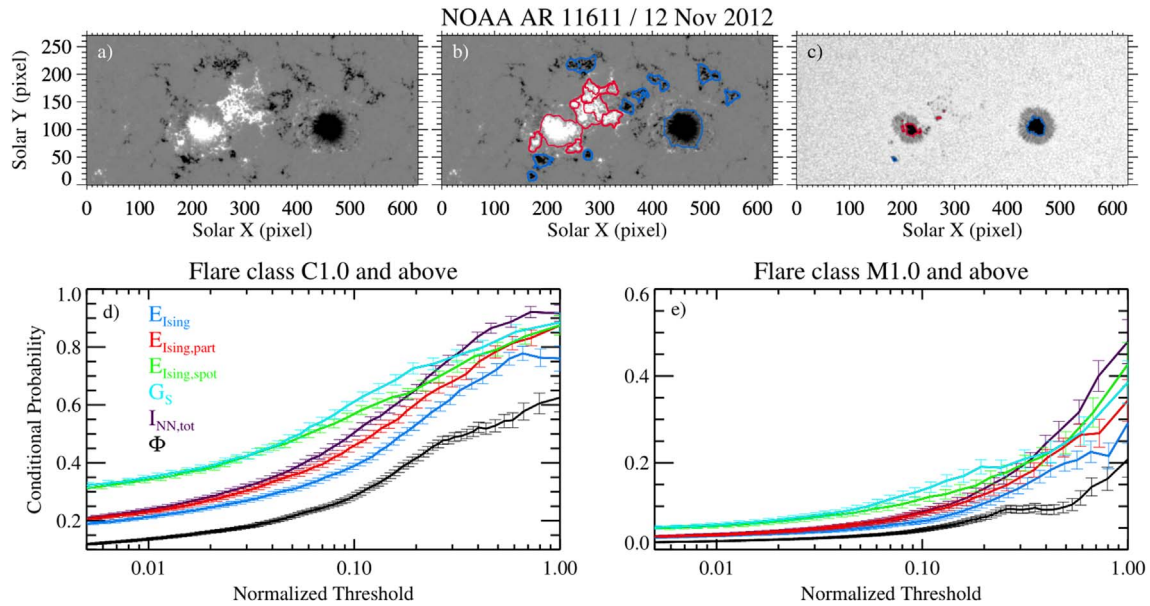


Fig. 10. Example of the development and tests of new morphological flare predictors (adapted from Kontogiannis et al., 2018). (a) Input LOS magnetogram of NOAA 11611. (b) The magnetogram is partitioned into positive- (red contours) and negative- (blue contours) polarity flux patches that are used to derive the Ising energy $E_{\text{Ising,part}}$ of the AR. These partitions, along with the vector magnetogram of the AR, are further used to calculate the total unsigned non-neutralized currents $I_{\text{NN,tot}}$. (c) Positive- (red) and negative- (blue) polarity umbrae identified in maps of continuum intensity are used to calculate the Ising energy of the umbrae $E_{\text{Ising,spot}}$ and the sum of the horizontal magnetic gradient G_s . (d–e) Conditional probabilities for GOES \geq C1.0 and \geq M1.0, respectively, calculated for a set of successive thresholds in E_{Ising} , $E_{\text{Ising,part}}$, $E_{\text{Ising,spot}}$, G_s and $I_{\text{NN,tot}}$. The respective probabilities inferred for the total unsigned magnetic flux Φ are also shown for reference.

for the forecast-observation pair arrays are resampled to create 2D index arrays of the form [“realizations”, “sample_size”]. For a bootstrap, indices in the range $[0, (N - 1)]$ are uniformly randomly drawn for a total of “realizations” \times “sample_size” times to directly populate the 2D index array. For a jackknife, each index is dropped to generate a 2D $[N, (N - 1)]$ index array. Finally, for a subsample, in each realization the monotonic array of indices $0, 1, 2, \dots, (N - 1)$ are randomly shuffled and the first “sample_size” number of indices are extracted and stored in the 2D [“realizations”, “sample_size”] index array. After the 2D index permutation sets are generated, the full suite of verification metrics and skill scores are calculated over the “sample_size” entries for each realization. These are achieved using `calc_class_stats()` and `calc_prob_stats()` once again to ensure consistency of the metric and skill score calculations with those reported as primary results in the verification data structure. To simplify the uncertainty assessment portion of the structure, any key names containing nested structures are removed (i.e., the contingency table in the classification data structure and the probability-thresholded reliability diagram in the probability data structure). Following this, averages and standard deviations are calculated over the number of realizations for each separate metric and skill score. The resulting dictionaries of average and standard deviation values are separately saved into the “unc_avg” and “unc_std” keys, respectively, in the output verification data structure.

We note for completeness that the above implementation of metric uncertainties is used because it is deemed as the most time and computationally effective. Alternative implementations can be pursued, but these are reserved for future FLARECAST infrastructure upgrades.

5 The FLARECAST science results

FLARECAST’s scientific outcome relied on the following four pillars: an investigation of new, or a revisit of potentially interesting, flare predictors (Sect. 5.1); the application of machine-learning methods to flare forecasting, also enabling feature (i.e., predictor) ranking in terms of predictor importance (Sect. 5.2); exploration of, first, eruptivity studies with synthetic, 3D MHD models (Sect. 5.3) and, second, the flare-CME connection (Sect. 5.4). This last component shows that, importantly, the FLARECAST AR property database can be meaningfully coupled with CME catalogs such as those of the EU HELCATS project¹⁶ or NASA’s DONKI¹⁷ to extend flare forecasting to eruptive flare (i.e., CME onset) forecasting with expectation values of CME properties. The complete list of FLARECAST-related peer-reviewed articles can be found in Appendix C.

5.1 New and revisited flare predictors

In spite of possessing one of the most comprehensive ensembles of flare predictors, FLARECAST opted to identify additional properties that could assist in flare forecasting. In this respect, subsets of the FLARECAST property (including flare association) database were used and linked to the flaring history of ARs. Promising quantities involved photospheric shear flows (Park et al., 2018), non-neutralized electric currents (Kontogiannis et al., 2017), magnetic gradients (Kontogiannis et al., 2018), and the differential emission measure calculated

¹⁶ <https://www.helcats-fp7.eu/>

¹⁷ <https://swc.gsfc.nasa.gov/main/donki>

above ARs (Gontikakis et al., 2020). In addition, FLARECAST has taken advantage of the results of Guerra et al. (2018), advocating for the joint use of line-of-sight and solar-surface-vertical versions of predictors in machine learning methods, as well as of the results of McCloskey et al. (2016), revisiting sunspot class associations with flaring rates and confirming the increased likelihood of flaring in case flux emergence enhances the photospheric compactness of ARs.

Strong shear flows are observed along intense magnetic PILs in the AR photosphere. PILs are arguably the single most telltale feature of enhanced flare productivity. Properties derived from a map of plasma shear flows were thus analysed by Park et al. (2018). In particular, an algorithm was developed to quantify photospheric shear flows in ARs in terms of three properties, treated as predictors in this framework: the mean ($\langle S \rangle$), maximum (S_{\max}) and integral (S_{sum}) shear-flow speeds along strong-gradient, strong-field PIL segments. This algorithm was then applied to a large data set of ~ 2500 co-aligned pairs of AR vector magnetograms with 12-min separation over the period 2012–2016. Park et al. (2018) found that ARs with more widespread and/or stronger shear flows tend to not only be more flare-productive over the next 24 h but also produce major flares sooner. This finding paves the way for future conditional predictions as well as forecasts better integrating the flare history of ARs. The importance of flare history in predictions of future flaring has already been identified by Falconer et al. (2012), as well as by comprehensive statistical studies (Leka et al., 2019a, b; Park et al., 2020). As further discussed in Park et al. (2018) and Welsch et al. (2009) shear-flow properties exhibit a weaker, albeit still positive, correlation with the flare peak flux than Schrijver's R . However, Park et al. (2018) revisited the problem using almost as large a sample of co-aligned magnetograms as Welsch et al. (2009), but at much higher quality (i.e., spatial resolution, cadence) and with a more detailed sample of NOAA-numbered ARs (i.e., in terms of observed times/locations, flaring activities and photospheric magnetic properties). Park et al. (2018) further focused on flaring activity over the next 24 h, as well as on the waiting time between major flares as a function of the shear-flow properties. We believe that the new results obtained with this improved magnetogram dataset and systematic study warrant additional investigation.

The predictive potential of properties related to non-neutralized electric currents in ARs was also investigated. Intense, shear-ridden PILs are exclusive areas of non-neutralized currents that imply the injection of net currents into the corona whose associated non-potential (i.e., free) magnetic energy could be instrumental for flares and eruptions (Török et al., 2014; Dalmasse et al., 2015). Georgoulis et al. (2012) had already proposed a method to calculate non-neutralized currents accurately, along with their applicable uncertainties. This method is slightly different than the one of Leka & Barnes (2003a) who added the absolute values of the total current from each magnetic polarity in that, first, it algebraically adds the total currents from both polarities to obtain the net current and, second, it performs a robust flux partitioning of a given magnetogram to identify and distinguish between magnetic polarities. The integral form of Ampere's law for the inference of net currents was also applied by Falconer et al. (2002); however, no specific discussion on the precise choice of contours was made. Relying on the method of Georgoulis et al. (2012)

and Kontogiannis et al. (2017) implemented the computation of the non-neutralized electric currents within the FLARECAST property database. Calculations for a sample of AR time series showed with statistical significance that the systematic increase in the amount of non-neutralized currents concurred with the development of strong PILs and signaled phases of intense flaring. Further application to a representative sample of Solar Cycle 24 ARs showed that the total unsigned non-neutralized current of an AR is promising in distinguishing between flaring and non-flaring AR populations within a 24-h time window (Kontogiannis et al., 2017). "Promise" in that work implied a significantly better (i.e., beyond applicable uncertainties) ability to distinguish between flaring and non-flaring active regions than the (often used as baseline) unsigned magnetic flux – see Figure 10 and discussions in Barnes & Leka (2008) and Georgoulis (2012, 2013).

Additionally, FLARECAST developed new algorithms to reproduce some recently proposed predictors not yet implemented in forecasting schemes, relying on the complexity of the spatial distribution of the photospheric magnetic flux. These were the sum of the horizontal magnetic gradient (c.f. Korsós et al., 2014, 2015; Korsós & Erdélyi, 2016) and the Ising energy (Ahmed et al., 2010). The sum of the horizontal magnetic gradient is calculated by combining magnetograms and photospheric continuum images. The latter are used to locate the umbrae of sunspot groups while the magnetograms are used to define the mean magnetic field strength in the umbrae. These values and the separation distances between opposite-polarity umbrae are used to calculate the sum of the magnetic field gradients between all possible opposite-polarity umbrae pairs. Furthermore, while the original formulation of Ising energy considers pairings between opposite polarities on a pixel-by-pixel level, FLARECAST examined that form as well as the Ising energy of opposite-polarity umbrae and topologically inferred magnetic partitions.

Kontogiannis et al. (2018) demonstrated that magnetic-gradient- and Ising-energy-based predictors are also worth including in automated flare forecasting schemes. In terms of Bayesian conditional flaring probabilities (Fig. 10), the sum of the horizontal magnetic gradient appeared to be the most efficient property, followed by the Ising energy of the umbrae pairs. Efficiency here implies again an ability to better distinguish between flaring and non-flaring active region populations, not necessarily in terms of predictive potential. The latter, of course, is to be tested in more strict terms within machine learning prediction methods employing multiple predictors and being able to rank them. Although many of these and other properties relate to PILs, Kontogiannis et al. (2018) showed that each of these properties exhibits specific dependence on the position of ARs on the solar disk and encompasses different information. These differences can justify the different efficiency of these flare predictors and their potential inclusion in multi-parameter machine learning-based forecasting schemes.

Gontikakis et al. (2020), in a forthcoming study, introduced a new, "orthogonal" view on flare prediction, analysing EUV image time series of ARs. More specifically, the temporal evolution of the DEM, calculated over ARs, was studied as a possible flare predictor. The DEM shows the distribution of the plasma EM as a function of temperature, T , and is derived from optically thin spectral lines. It is defined as $\text{DEM}(T) = n_e^2 dl/dT$ and involves the derivative of the line-of-sight l over

the plasma temperature, with n_e being the electron number density. Syntelis et al. (2016) found that the DEM calculated over an intensely eruptive AR increased substantially a few (i.e., ~5) hours before one of the eruptions, with the other eruption following shortly after. This was interpreted as a possible pre-flare manifestation (see, e.g., Fletcher et al., 2011). To achieve statistical significance in a DEM-based pre-flare activity study, Gontikakis et al. (2020) used DEM inferences by the GAIA-DEM archive.¹⁸ In GAIA-DEM, the DEM is assumed to be a Gaussian function of $\log(T)$ (Guennou et al., 2012) and the archive provides images of three Gaussian parameters, along with χ^2 . These parameters are the EM (i.e., the integral of the DEM), the temperature T_{\max} where the DEM is maximized, and the Gaussian width of the DEM. Gontikakis et al. (2020) used 6-h long time series of each DEM parameter. The DEM parameters were computed from solar structures extracted from 9000 HMI HARPs. Positive derivatives of the time series, indicating that AR plasma heating is at work, were treated as indicators of imminent flares. Analyzing the DEM time series appeared to give more significant conditional flare probabilities than the reference unsigned magnetic flux (i.e., its value at the end of the time series) for GOES \geq M1.0 flares, but it was less successful for GOES \geq C1.0 flares. These may be grounds to advocate that, provided that the significant uncertainties in the DEM calculation can be constrained, DEM time series may complement or even enhance the short-term predictive ability of photospheric properties, at least for major flares.

5.2 Flare prediction and feature ranking

The FLARECAST machine-learning prediction component draws from and relies on the plethora of predictors extracted from NRT HMI SHARP magnetograms populating the project's property database. This central objective has been realized by Benvenuto et al. (2018), Florios et al. (2018), Piana et al. (2019) and Campi et al. (2019), while the overall process has been described and discussed by Massone et al. (2018).

Benvenuto et al. (2018) tested an array of supervised and hybrid (i.e., comprising both supervised and unsupervised elements; see Sect. 2.3) machine-learning methods on historical NOAA/SWPC sunspot classifications between August 1996 and December 2010, training the methods on similar data between December 1988 and June 1996. They relied on flaring rates associated to certain sunspot classes, similar to McCloskey et al. (2016), but using machine learning for prediction. The main result was that hybrid methods tend to outperform strictly supervised ones and approach the performance of clustering methods. Benvenuto et al. (2018) used a variety of metrics and skill scores present in Table 4 and further determined that a reliable feature ranking by means of their prediction value is possible.

Florios et al. (2018) used three different supervised methods (i.e., MLPs, SVMs and RFs; see Table 2) on a subset of the FLARECAST property database over a five-year interval covering 2012–2016. They were able to correlate between the different methods and concluded that RFs (Breiman, 2001) could be

the method of choice in a routine, operational flare forecasting scheme. Feature ranking was performed by means of different scores and indices and it was found that peak performance could be achieved by using 10–12 predictors in total. Both purely probabilistic and probability-thresholded categorical performance verification was performed, with optimal probability thresholds determined. The overall performance practically matched, and in some cases exceeded, that of Bobra & Couvidat (2015) who used a SVM and performed the first flare forecasting work using HMI SHARPs.

The study of Campi et al. (2019) performed a data- and property-intensive flare prediction investigation. It utilized a database of 14,931 point-in-time property vectors, each comprising up to 171 predictors within the interval September 2012 – April 2016. The main objective was to determine whether, and to what extent, this abundance of predictors is essential or redundant in flare prediction. The study utilized two machine-learning methods, namely hybrid LASSO and RFs, and opted to construct four training and four testing sets, each corresponding to a 24-h forecast window from a specific issuing time (i.e., 00:00, 06:00, 12:00, and 18:00 UT each day). Particular attention was paid to the complete separation between training and testing sets, that were not only non-time-overlapping (i.e., results from each issuing time were not combined) but they used different ARs (i.e., HARP numbers, in implementation) for training and testing. As such, another objective was to find the impact of different intervals of the *same* ARs being used in both training and testing. The main findings of Campi et al. (2019) were the following:

- Properties with the best predictor ranking have the smallest variance in ranking, which implies that their impact on prediction is consistently high. However, the order of best-ranking predictors varies for different prediction methods.
- Particularly for prediction of GOES \geq C1.0 flares, the best-ranking predictors are common across all issuing times studied. This is less of a case for the prediction of major flares of GOES \geq M1.0 flares.
- Only a small number of predictors (different for each method, however, most notably for major flares) suffices to make the prediction method achieve its top performance. Hence, the vast majority of predictors are redundant in such approaches and, in fact, using all of them can even be detrimental to performance. The redundancy of predictors was also noted in Barnes et al. (2016) as possibly due to intrinsic, but not necessarily known ab initio, correlations between different predictor values. This effect may also be enhanced for events of increased rarity and hence poorer statistics, i.e., when forecasting increasingly higher flare classes.
- The stricter approach of separating ARs between training and testing sets leads to notably lower skill score values. Relaxing the robustness of training and testing by allowing “information leaking”, not necessarily in terms of temporal overlapping but in terms of different intervals in the same ARs, readily increases these values. Improved performance in this case, however, is due to memorization, rather than learning, on behalf of the machine learning method used. Lower skill scores for unbiased training and testing, as also found by Piana et al. (2019), reflect the intrinsic, uncompromising stochasticity of the flaring phenomenon.

¹⁸ The GAIA-DEM is a MEDOC database that archives full-disk maps of DEM parameters, derived from the 6 EUV channels of the SDO/AIA telescope.

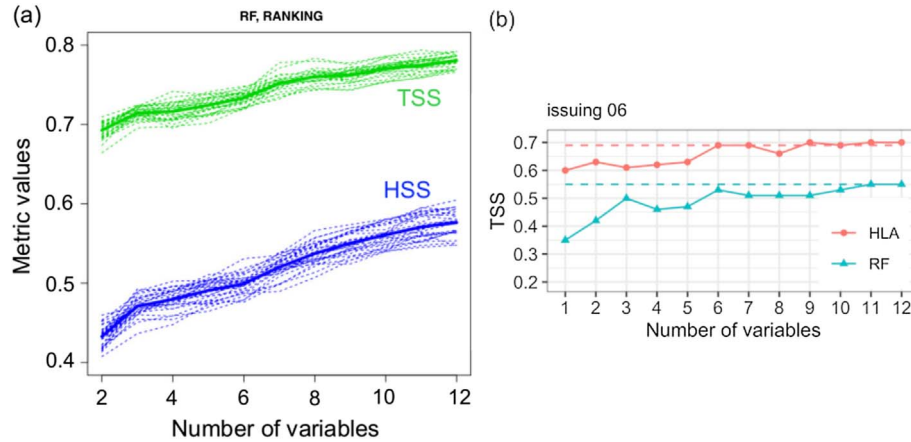


Fig. 11. Example prediction runs adapted from (a) Florios et al. (2018) and (b) Campi et al. (2019) for the same basic prediction settings, namely a 24-hr forecast window for GOES \geq M1.0 flares. Both TSS and HSS values are shown in *a* using RFs only, while *b* shows only TSS values for two different methods, namely hybrid LASSO (HLA) and RF. The abscissas in both plots correspond to the number of predictors used.

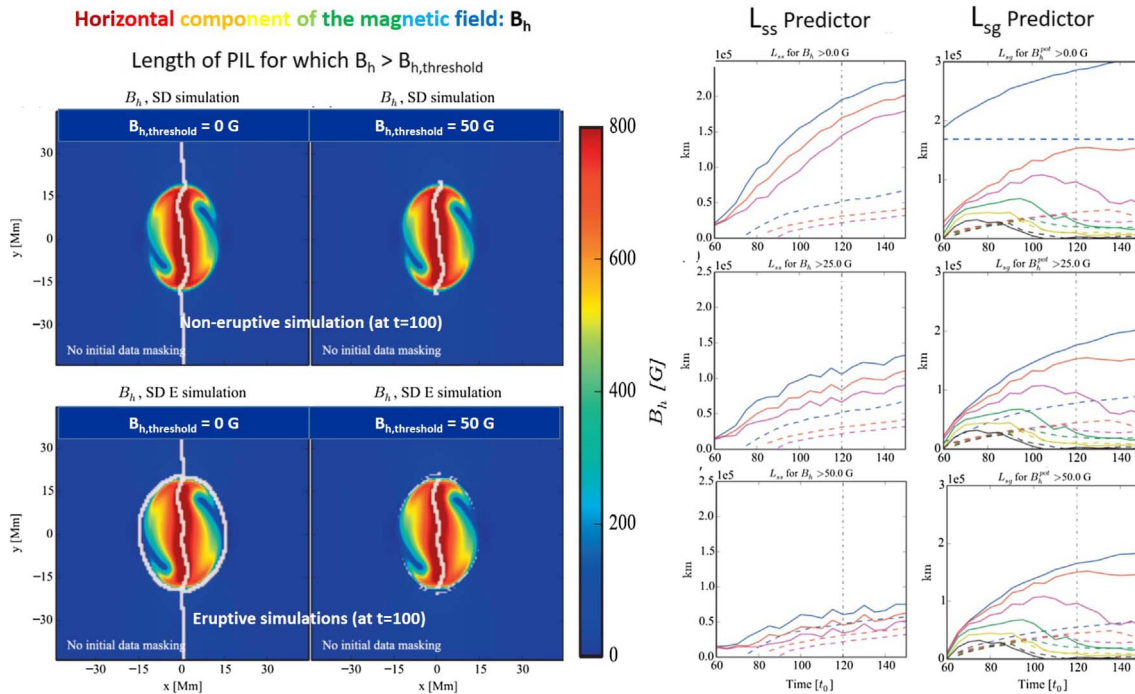


Fig. 12. Example of the impact of thresholding, here of the threshold $B_{h,th}$ for the horizontal magnetic field strength B_h , on the efficiency of two potential flare predictors, namely the lengths of the strong-shear and strong-gradient PIL, L_{ss} and L_{sg} , respectively (see text for details). Figure adapted from Guennou et al. (2017).

We note at this point that the relatively small number of best predictors found by Florios et al. (2018) and Campi et al. (2019) essentially aligns with earlier results by Ahmed et al. (2013), Bobra & Couvidat (2015) and Bames et al. (2017). The latter study also found that a set of three or four predictors was sufficient to guarantee best performance for two classifiers, NPDA and RF. That study addressed prediction of both filament eruptions and flares and found significant overlapping of about half of best predictors between the two types of predictions. In the case of Campi et al. (2019), however, such predictor patterns

were found only for the prediction of GOES \geq C1.0 flares within different classifiers, while mostly different sets for different classifiers were found in the prediction of major flares (i.e., GOES \geq M1.0).

Both Florios et al. (2018) and Campi et al. (2019) used various metrics and skill scores from Table 4 but focused on HSS and TSS. Example runs by both studies and their associated HSS and TSS values are shown in Figure 11. This leads to the conclusion that, for both studies, the metric values increase while adding more predictors, although in both cases the

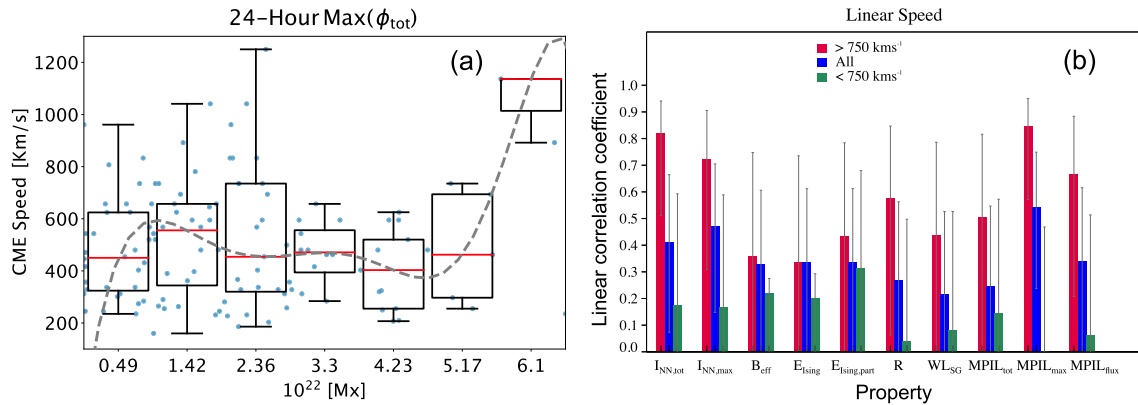


Fig. 13. (a) Correlation between CME speed and the peak of total, unsigned magnetic flux in the source active region up to 24 h prior to the eruption. Points correspond to the actual data values for 124 CME – source region pairs, while a box-and-whiskers plot is overlaid along with a spline fit to quantify the correlation. Reproduced from Murray et al. (2018). (b) Linear correlation coefficient and respective uncertainties between source region properties and CME speed for a sample of 32 CME – source region pairs. A distinction between fast (red columns) and slow (green columns) CMEs has been made while results for the entire CME sample are shown in blue columns. Reproduced from Kontogiannis et al. (2019).

amplitude of the increase and the values achieved depend on the adopted skill score. Furthermore, in Florios et al. (2018) the increase seems to persist after 12 predictors while in Campi et al. (2019) saturation is reached with just six predictors (although when RF is implemented on three predictors, it leads to a TSS value very close to the one obtained with six predictors). The difference between the two studies (and subsequent panels in Fig. 11) is probably due to the different way the training set is populated, the latter experiment not mixing up active regions between the training and test sets. We notice, moreover, that the TSS value for Campi et al. (2019) is lower than that of Florios et al. (2018) for both applied methods, despite the fact that the predictors used in Florios et al. (2018) are a subset of those used in Campi et al. (2019). This is again a result of the more stringently separated training and testing procedure followed by Campi et al. (2019). In fact, the prediction method itself may often be less important than the robustness of training and testing practices. By roughly reproducing the training and testing of Bobra & Couvidat (2015), Florios et al. (2018), Campi et al. (2019) were able to practically match the performances of these studies for different machine learning methods. As a result, the “uncompromised” stochasticity of flare occurrence, as coined by Campi et al. (2019), could not be curbed even with one of the largest-dimensionality parameter spaces.

Given the above, the FLARECAST performance verification infrastructure (Sect. 4) is in place and allows the evaluation of performance of any future effort than can rely on either the existing or enhanced property database (Sect. 2.2) and the prediction database (Sect. 2.3). Obviously, the quest continues.

At this point it is important to clarify that, although FLARECAST valued highly the use of predictor time series for flare forecasting, this important part of the analysis did not have the time to materialize during the project’s nominal period of execution. The only exception is the forthcoming work of Gontikakis et al. (2020) that exploits DEM time series to uncover flare predictors or precursors. Time series, in terms of flare history and/or predictor evolution, have been found useful

in flare prediction (Leka et al., 2018, 2019b; Park et al., 2020), going beyond point-in-time forecasting. While this avenue is yet to be investigated with the FLARECAST databases, these data clearly enable one to pursue this objective in the future.

5.3 Studies using synthetic data

While observational data are the ultimate reference frame to determine the flare-predictive ability of a given quantity, these data have critical limitations toward a complete and pertinent description of solar ARs. For example, while it is agreed that the 3D coronal distribution of the AR magnetic field plays a key role in the triggering and development of instabilities leading to flares and eruptions in general, only the 2D photospheric field vector can be routinely inverted from Stokes images by means of the Zeeman effect. Thanks to tremendous developments in computational solar physics, numerical experiments of flare and eruption triggering are becoming increasingly common and realistic (see, e.g., the review of Green et al., 2018). These experiments have now made it possible to explore new flare predictors. Kusano et al. (2012) have carried out numerous 3D MHD simulations in which the initial set-up was parametrically modified to accommodate eruptive and non-eruptive configurations. Distinguishing between flaring and non-flaring configurations is essential in investigating the efficiency of potential predictors.

This exploratory component of FLARECAST was facilitated by the 3D MHD simulations of Zuccarello et al. (2018), who extended upon Kusano et al. (2012). These simulations started with a dipole magnetic configuration, emulating the two polarities of an AR, and were driven by specific, carefully modeled and controllable photospheric velocity flows. Four types of converging flows were tested (all eventually leading to eruptions) against a control, non-eruptive case. Depending on these driving flows and their experimental control, eruptions took place at different times that, however, could be determined precisely for each simulation. Given the full 3D data sets, different scalar volumetric quantities were calculated, among which were the potential and non-potential (i.e., free) magnetic energy

budgets and the relative magnetic helicity H , that was further decomposed into non-potential and volume-threading helicity terms (c.f. Berger, 2003; Moraitis et al., 2014; Linan et al., 2018). At the onset of eruptions, all parameter values showed a significant dispersion, indicating the lack of a practical instability threshold. Interestingly, no particular free-energy threshold could be identified for flares and eruptions in general, despite the key role of the magnetic free energy in eruptions. However, the ratio of a specific term H_j (coined the “non-potential” helicity) to the total relative helicity behaved differently; eruptions seemed to always take place for the same value of the ratio H_j/H . Therefore, this ratio may represent a critical threshold at which a modeled AR becomes unstable.

The importance of the ratio H_j/H was further highlighted by Pariat et al. (2017) who analysed parametric 3D MHD simulations of the formation of solar-like ARs. These simulations were originally presented by Leake et al. (2013, 2014) and modeled the self-consistent formation of ARs via the emergence of a twisted magnetic flux tube from the upper convection zone into the solar atmosphere. Flux emergence led to the formation of a magnetic configuration similar to standard bipolar ARs. The parametrisation of the initial conditions enabled the simulation of both quiescent and eruptive configurations. In Pariat et al. (2017), the full 3D MHD simulated coronal magnetic field for different cases was investigated and analyzed, using magnetic energy and helicity. While magnetic energy, in particular free magnetic energy, exhibited strong differences during the evolution of various configurations, some helicity-based quantities were robustly discriminating between eruptive and non-eruptive simulations. The total relative magnetic helicity H was not one of them, showing its largest budget for the non-eruptive simulation. However, the ratio H_j/H showed high values for eruptive simulations only, most importantly *before* the onset of eruptions. Shortly after eruptions, its values were similar for both eruptive and non-eruptive simulations. As a result, the ratio H_j/H was portrayed as a promising eruptivity proxy.

Given the above, the actual predictive value of the ratio H_j/H will ultimately be determined by data-driven modeling of observed solar ARs. If future computational advances make the meaningful calculation of the non-potential 3D coronal magnetic field feasible in NRT, testing of this proxy will become possible and may lead to innovative flare predictors in future, advanced versions of FLARECAST and other operationally-oriented systems.

Synthetic data obtained from numerical simulations can also be used to study flare proxies following the same methodology with observed data. Guennou et al. (2017) studied flux emergence simulations by Leake et al. (2013, 2014) in a way that was similar to the treatment of observational data, namely by isolating the 2D magnetic field vector of the modeled AR photosphere. A list of about 100 different scalar properties was then computed in a way similar to the procedures developed in FLARECAST. Guennou et al. (2017) found that, among the properties tested, only those associated with PIL features presented significant preflare signatures. Virtually all other properties were not able to differentiate between eruptive and quiescent simulations. Since the computation of several of these properties depends on the choice of parameters (e.g., thresholds for masking the data, magnetic field thresholds, etc.), Guennou et al. (2017) could study the role of these thresholds on the

ability of a property to determine the eruptivity. As exemplified in Figure 12, it was found that some properties were highly sensitive to the choice of user-defined thresholds whereas some others were more robust. The four panels on the left show the distribution of B_h at the modeled photospheric boundary of two 3D MHD flux emergence simulations at $t = 100$, in normalized time units of the non-eruptive (top) and eruptive (bottom) simulations. The white curves show the portion of the PIL where $B_h > B_{h,th}$, for two different choices of $B_{h,th}$ (0 G and 50 G). The length of the PIL where $B_h > B_{h,th}$ is one of two criteria used in the computation of L_{ss} , while L_{sg} uses a different criterion based on the potential field, namely the length of the PIL where $B_{h,pot} > B_{h,th}$.

The six plots on the right show the temporal evolution of L_{ss} and L_{sg} for the eruptive (solid curves) and non-eruptive (dashed curves) simulations under three different values of $B_{h,th}$, namely 0 G (top), 25 G (middle), and 50 G (bottom). The different colors corresponds to different choices of criteria also entering in the computation of L_{ss} and L_{sg} , with details discussed in Guennou et al. (2017). As $B_{h,th}$ increases, L_{ss} becomes similar for the non-eruptive and the eruptive simulation. This predictor is thus highly sensitive to the choice of $B_{h,th}$. Conversely, L_{sg} is relatively insensitive to the choice of $B_{h,th}$. As a result, L_{sg} is capable of discriminating the two simulations and could conceivably be used as an eruptivity proxy.

In addition, Guennou et al. (2017) studied the robustness of certain parameters in the presence of noise, reporting that parameters related to electric currents seem to be the most sensitive. This would imply the need for screening such predictors before inserting them in prediction pipelines.

5.4 Flare-CME connection studies

The FLARECAST exploratory component further enabled researchers to expand upon the project results into other cutting-edge research areas, including coronal mass ejection (CME) prediction. Space-weather operations are currently lacking sufficient warning for eruptive flares, with forecasters only predicting CME arrival at Earth once an eruption is observed (see, e.g., Möstl et al., 2014; Verbeke et al., 2019). One may reasonably expect that CMEs and other solar eruptive phenomena can be physically linked by combining data from a multitude of ground- and space-based instruments, as well as models. However, this can be challenging for automated operational systems. Understanding the processes and sources involved in solar eruptions that could enable the use of photospheric properties is imperative for improving this area of forecasting.

The flare-CME connection that could advance CME prediction was investigated via a unique synergy between FLARECAST and an EU Framework Programme 7 project funded under the “*Exploitation of space science and exploration data*” Call. The HELCATS project created numerous data products from heliospheric imaging onboard the two NASA STEREO spacecraft in order to track the evolution of CMEs in the inner heliosphere. Using the project’s main catalogue of over 2000 CME events imaged between 2007 and 2017, an automated algorithm was developed to connect the CMEs observed by STEREO to any corresponding solar flares and AR sources on the solar surface, resulting in the HELCATS LOWCAT

catalogue. Murray et al. (2018) supplemented the LOWCAT catalogue's information on CME kinematic properties (such as speed and angular width) with FLARECAST's extensive property database derived from vector magnetograms, enabling a deeper study into the characteristics of eruptive ARs. Initial statistical analysis was undertaken on the new combined data set, with total unsigned flux, vertical current density and current helicity identified as properties of interest for potential eruption warning thresholds. Beyond the scientific gain, this collective effort provided an excellent opportunity to foster communication between different, large-scale research projects which may work in parallel on similar research topics to benefit space-weather forecasting. See Figure 13a for an example of this synergy, while further details can be found in Murray et al. (2018).

A further detailed investigation was performed for a smaller sample of 32 eruptive flares (4 C-class; 16 M-class; 12 X-class) by Kontogiannis et al. (2019), seeking correlations between selected AR non-potentiality properties and CME kinematic characteristics. The selection intentionally focused on PIL-related properties, including some well-established ones (such as Schrijver's R , WL_{SG} and B_{eff}) along with ones implemented in the course of FLARECAST (such as the Ising energy and the total non-neutralized electric current; see Table 1). CME information was collected from NASA's DONKI database, utilizing CME images by LASCO onboard the SOHO mission. Positive correlations between CME kinematic properties were found for all pre-eruption photospheric properties, in line with the results of Murray et al. (2018). Remarkably, correlations were stronger for faster CMEs implying that selected non-potentiality properties may be indicative of upper-limit eruption scales that ARs can produce over a given timescale. Among the FLARECAST properties tested, the preflare total unsigned non-neutralized current and the length of the main PIL stand out, exhibiting stronger correlations with the CME kinetic energy and speed (Fig. 13b). This finding points to a causal relationship between non-neutralized electric currents (stemming from compact, sheared PILs) and the associated free magnetic energy and helicity budgets in ARs and CMEs. Therefore, one envisions a future forecasting capability of basic CME kinematic properties *before* CMEs actually occur that, combined with the potential CME launch location as per its source AR, could give educated guesses of possible CME shock formation and geoeffectiveness.

6 Conclusions

FLARECAST was a multifaceted, interdisciplinary project. As such, it is appropriate to identify the different activity areas of the project and summarize conclusions pertinent to each of them.

6.1 Science and technology

One of the project's main aims was to seek the most efficient ways of forecasting solar flares. In this respect, FLARECAST is arguably the most intensive and systematic solar flare forecasting effort to date due to (i) the amount, cadence and validation of the input SDO/HMI SHARP magnetogram

data, (ii) the sheer number of AR properties treated as flare predictors (209 of them, with 171 used in the comprehensive study of Campi et al., 2019), (iii) the breadth of machine-learning prediction algorithms implemented (14 included in Table 2), and (iv) the comprehensive performance verification approach and infrastructure, along with suggestions on how to implement it in a rigorous training and testing philosophy.

Solar flares are long known to be stochastic events, namely responses of strongly nonlinear (quite likely self-organized critical) dynamical systems, as solar ARs are thought to be. Hence, flare forecasting appears intrinsically probabilistic, notwithstanding that some machine-learning prediction methods provide a binary output. Despite this long-standing knowledge, the FLARECAST Consortium was keen to determine whether a combined Big Data and machine learning effort could lift this stochasticity barrier by achieving a compelling, verified binary separation between flaring and non-flaring AR populations. As Campi et al. (2019) declared right from that paper's title, and was also implied by the results and discussion of Florios et al. (2018), stochasticity in flare occurrence and subsequent forecasting remains uncompromising. This, of course, could be due to limitations or shortcomings of the tested methods and/or data. Regardless, despite significant progress brought by FLARECAST and many other efforts worldwide, a decisive (i.e., binary and with high enough verification metric values to be used for decision making on an operational context) flare prediction capability is still beyond reach. This said, FLARECAST has generated codes, data and infrastructure that could potentially facilitate this crucial milestone in the future.

A recent trend in the literature (see, e.g., Nishizuka et al., 2018; Li et al., 2020, and follow-up works) suggests that machine learning may not even be sufficient, so the community should be moving toward deep learning flare prediction efforts. This remains to be seen; even in the above deep learning efforts, however, a significant degree of stochasticity persists, reflected in less-than-optimal performance indicators. So far, solar deep-learning prediction methods have by no means outperformed machine learning ones. It is an open question whether this can be reversed in the future, although the massive data needs of deep learning algorithms (e.g., Goodfellow et al., 2016) may render flare prediction a less-than-optimal domain for deep learning efforts due to the lack of sufficient event instances to train on.

In terms of performance verification, while FLARECAST created a detailed apparatus (Sect. 4), the nominal duration of the project did not allow its complete exploitation. Existing project studies (Benvenuto et al., 2018; Florios et al., 2018; Campi et al., 2019) have used meaningful parts of it but not the entire infrastructure provided in Table 4. Future studies, however, will have the option to work and rely on this infrastructure for a comprehensive performance verification assessment.

While FLARECAST was a case study of R2O, it soon became clear that we could learn from its output, by means of both predictor ranking and the explorative research component. Then, part of the project's output contributes to O2R objectives. In particular, a robust ranking of predictors triggers the question of why just a few ($\lesssim 10$) predictors, different for each prediction method at least for flares of GOES class $\geq M1.0$, gather virtually all the predictive power in Campi et al. (2019), while in another attempt by Florios et al. (2018) that used the same (RF) method the increase in performance seems monotonic as more predictors are added. The above generalize and diversify the

big picture that now calls for (i) a comprehensive review of training and testing practices, regardless of prediction methods, and (ii) an investigation of specific predictors' performance on given prediction method(s). In light of all these, one key take-away message is that we should not be employing "black box" (i.e., non-human-interpretable) prediction methods, in spite of their performance, in order to keep the knowledge discovery channel open (Rudin, 2019).

In terms of exploration, the project used synthetic data (Guennou et al., 2017; Pariat et al., 2017) to investigate elements not accessible by the available observational data, but also concerned itself with the flare-CME connection to investigate flare predictors exhibiting a statistically significant correlation with CME parameters. This was pursued via both the FLARECAST – HELCATS synergy (Murray et al., 2018) and the FLARECAST – DONKI association of Kontogiannis et al. (2019). These works, along with previous ones in this respect (Falconer et al., 2002; Barnes, 2007; Ugarte-Urra et al., 2007; Georgoulis, 2008), all show promise toward a viable interpretation of the pre-eruption configuration in solar ARs and a deeper, more complete understanding of solar eruptions.

Along with other recent studies (Barnes et al., 2016; Leka et al., 2019a, b; Park et al., 2020), FLARECAST has showcased the need for fixed, pre-defined data sets in solar flare (and eruption) forecasting. Such "benchmark" data sets (see Angryk et al., 2020, for a definition and an example) enable the precise evaluation of different methods, machine/deep learning or other, on preset training and testing samples. Performance comparisons on different data sets and/or phases of the solar cycle can be problematic due to the differences in underlying flare statistics over different time periods, so a benchmark dataset should be partitioned in such a way as to treat variable climatology to the extent possible. The FLARECAST property database could also play the role of a benchmark dataset. It further enables the important action of time series forecasting, that was unfortunately not achieved during the project's nominal period of execution, and the treatment of class imbalance in machine learning performance. Such investigations directly address another trait of major flares, namely their scarcity and rareness, that becomes extreme for flares of historic magnitudes.

FLARECAST developed an infrastructure that could help address the need for an integrated space weather forecasting platform. Coordination of space weather forecasting efforts has been identified as a key priority in proposed roadmaps such as the one developed on behalf of the COSPAR (Schrijver et al., 2015). It is also prominently recommended by the ESWACC (Opgenoorth et al., 2019). In line with these studies, global coordination of space-weather forecasting efforts sprung out as a top-level conclusion during the FLARECAST Stakeholder Workshops. The project's infrastructure could serve as a testbed, or breadboard, for future applications encompassing the other two main legs of "stormy" space weather, namely CMEs and SEPs. Given the versatility of the Docker engine and containers used in FLARECAST, components of different programming languages can be removed, updated, or added in a highly modular fashion.

All in all, FLARECAST has amply shown that solar flare forecasting should not be, and thankfully is not, an internal affair for heliophysicists. Fusion of expertise is paramount to advance

and, ultimately, break ground toward an efficient space-weather forecasting but also other complex real-world problems.

6.2 Operations

By defining different flare forecasting modes (i.e., based on GOES flare class, forecast window, latency, etc.) and testing numerous prediction algorithms, FLARECAST reaffirmed that one single recipe does not fit every need. Patterns between different prediction methods were seen in the best-performing predictors for GOES \geq C1.0 flares (i.e., flares of GOES C-class and above), for example, but not for GOES \geq M1.0 flares (Campi et al., 2019). In addition, AR properties that correlate best with CME properties (Murray et al., 2018; Kontogiannis et al., 2019) may, or may not, be within the best performing predictors for a given method. We view these findings in light of an outcome of the FLARECAST Stakeholders' Workshops that the real end users of the project's results are most likely expert operational forecasters who can disseminate them as needed. Other synergistic community works fall along similar lines by arguing for the benefit of human "forecasters in the loop" (Leka et al., 2019b). It then becomes clear that more time and effort will be needed before agencies and institutions committed to operations are able to shift from today's relatively simple forecast philosophies to sophisticated, multi-dimensional parameter spaces, Big Data and intensive machine/deep learning methodologies. It is a crucial step to be taken, however, balancing on a fine line between simplicity and interpretability of the concepts and the necessary level of complexity dictated by the problem at hand.

In brief, in spite of its limited duration that left a few envisioned elements to be realised (i.e., time series forecasting and comprehensive performance verification of all prediction algorithms), we hope that FLARECAST can be thought of as a future "textbook" R2O case, paving new ways and even narrowing "the valley of death" of less-than-successful approaches between the two domains of research and operations. As eloquently put by Merceret et al. (2013), who drew analogues between space weather and terrestrial meteorology, there may be leads able to transform the "valley of death" into a "valley of opportunity".

It should not go unnoticed that the immediate next step for the FLARECAST infrastructure is to become a flare prediction service of ESA's SSA Space Weather Service Network¹⁹. This task is undertaken by the FHNW partner that has originally developed the project's technology apparatus.

6.3 Communication

A key objective of the project was the communication with government and industry representatives, that cultivated into two FLARECAST Stakeholders' Workshops. A top-level conclusion of them was that communities involved in (solar flare or space-weather) forecasting should be in close contact. There are "language" and terminology barriers that require a continual effort toward keeping the momentum of communication and better defining the constantly evolving landscape of user needs. A prominent example in this direction is the aviation and power industries that have moved beyond colloquial

¹⁹ Available at <https://swe.ssa.esa.int/ssa-space-weather-activities>.

interest and are raising their awareness toward the actual capabilities, shortcomings and detailed products of operational forecasting.

We found from these Stakeholders' Workshops that accuracy, timeliness, and ease of use are the most important factors in a forecast from the users' perspective. Less important is the scientific detail accompanying each forecast. As mentioned above, it was also found that there should be an intermediate step between the researchers who devise the forecasts and the end users. This step is filled by operational forecasters, who can interpret forecasts and convey their messages in a language that users can understand and cope with.

As to the communication with the wider public, FLARECAST reaffirmed that the public is keen to listen and eagerly seeks knowledge on solar flares and space weather. It became clear that open-access, public-domain projects such as this one should return part of their discovered knowledge to the taxpayer citizen, by educating them credibly and responsibly, steering clear from the (unfortunately high) level of misinformation spread largely untested and un-scrutinized over the internet and social media²⁰.

6.4 Lessons learned

In implementing a sizable and diverse project such as FLARECAST, there was no shortage of lessons to learn. The intrinsic complexity of the project was not only due to its diversity, but also because Consortium partners were geographically distributed throughout Europe. This required certain steps to be taken in order to secure efficient coordination and implementation. Even so, however, some desired tasks could not be implemented within the nominal duration due to lack of time. The most important lessons were the following:

- An effective project governance is instrumental for planning ahead. Well-defined, distinct roles foreseen in the Consortium Agreement such as those of the Project Coordinator, Project Scientist, Financial and Legal Manager, as well as bodies such as the Project Management Board and the Steering Committee, provide a governance structure that can, in principle, lead the project into fruition via a properly shared workload. There should be space for some adjustments in the initial plan. However, radical changes midway through the project will only lead to confusion and impede implementation.
- The Consortium further needed an efficient way of remote communication, beyond e-mails, phone calls and teleconferences. This was achieved via an integrated platform of collaborative software tools that were utilized on a Consortium-wide level. This system combined flexible internal wiki-pages (e.g., enabling tagging individuals in documents with associated email updates, real-time recording of meeting minutes, efficient exporting of periodic report documents and code documentation, etc.) with integrated code repositories that have easily interpretable visualization

interfaces for version tracking (hence, allowing efficient code reviews and prompt troubleshooting of change-related issues).

- In Consortium deliberations, common understanding and a continual pursuit of consensus in a collegial atmosphere were valued as the most important elements for a successful collaboration.
- In a project with strict requirements for hardware installation and software development, time (and backup time) is necessary; efficient time management is paramount for the aversion of implementation delays. Given the high degree of synchronization required between parallel tasks, delays often seem inevitable; when they happen, they may also cause a domino effect affecting different project parts. Advance risk assessment and time planning help mitigate the impact of these delays.

Acknowledgements. The FLARECAST project was funded by the European Union Research and Innovation Programme under Grant Agreement no. 640216. The project would not have been possible without NASA's SDO mission and the unrestricted access to SDO data, primarily from Stanford University's JSOC. While we wish to sincerely thank all community members who participated in the public discussions, debates and the Stakeholders' Workshops, the FLARECAST Consortium is particularly grateful to the EU-assigned Project Reviewer, Thierry Dudok de Wit, and the EU Project Officer, Sabri Mekaoui, for their invaluable and continuous help and support. The project's administrative component has benefited enormously by the administrative structures of all partners, coordinated flawlessly by the Project Manager, Vangelis Argoudelis. We are also indebted to the FLARECAST Steering Committee (Neal Hurlburt [Chair], Graham Barnes, Douglas Biesecker, Pedro Russo, and Silvia Villa) who devoted time and effort on the project and contributed decisively to its smooth implementation. Finally, we deeply appreciate the contribution of the two anonymous reviewers whose insightful comments helped us substantially improve the presentation, content and context of the original manuscript. The editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

- Abramenko VI. 2005. Multifractal analysis of solar magnetograms. *Sol Phys* **228**(1–2): 29–42. <https://doi.org/10.1007/s11207-005-3525-9>.
- Abramenko VI, Yurchyshyn VB, Wang H, Spirock TJ, Goode PR. 2003. Signature of an Avalanche in solar flares as measured by photospheric magnetic fields. *Astrophys J* **597**(2): 1135–1144. <https://doi.org/10.1086/378492>.
- Ahmadzadeh A, Aydin B, Georgoulis MK, Kempton DJ, Mahajan SS, Angryk RA. 2021. How to train your flare prediction model: revisiting robust sampling of rare events. *arXiv e-prints* arXiv:2103.07542.
- Ahmed O, Qahwaji R, Colak T, Dudok De Wit T, Ipson S. 2010. A new technique for the calculation and 3D visualisation of magnetic complexities on solar satellite images. *Vis Comp* **26**: 385–395. <https://doi.org/10.1007/s00371-010-0418-1>.

²⁰ As a side note, recent efforts to counter the spread of fake news in the internet also rely heavily on the use of (interpretable) deep learning architectures (e.g., Shu et al., 2019).

- Ahmed OW, Qahwaji R, Colak T, Higgins PA, Gallagher PT, Bloomfield DS. 2013. Solar flare prediction using advanced feature extraction, machine learning, and feature selection. *Sol Phys* **283**: 157–175. <https://doi.org/10.1007/s11207-011-9896-1>.
- Al-Ghraibah A, Boucheron LE, McAteer RTJ. 2015. An automated classification approach to ranking photospheric proxies of magnetic energy build-up. *579*: A64. <https://doi.org/10.1051/0004-6361/201525978>.
- Albeverio S, Jentsch V, Kantz H (Eds). 2006. *Extreme events in nature and society*, Springer, Berlin Heidelberg. <https://doi.org/10.1007/3-540-28611-X>.
- Anastasiadis A, Papaioannou A, Sandberg I, Georgoulis M, Tziotziou K, Kouloumvakos A, Jiggins P. 2017. Predicting flares and solar energetic particle events: The FORSPEF tool. *Sol Phys* **292**: 134. <https://doi.org/10.1007/s11207-017-1163-7>.
- Anderberg MR. 2014. *Cluster analysis for applications: Probability and mathematical statistics: A series of monographs and textbooks*, Vol. 19, Academic Press, London. ISBN 978-0120576500.
- Angryk RA, Martens PC, Aydin B, Kempton D, Mahajan SS, et al. 2020. Multivariate time series dataset for space weather data analytics. *Nat. Sci. Data* **7**(1): 227. <https://doi.org/10.1038/s41597-020-0548-x>.
- Archontis V, Török T. 2008. Eruption of magnetic flux ropes during flux emergence. *A&A* **492**(2): L35–L38. <https://doi.org/10.1051/0004-6361:200811131>.
- Aschwanden MJ, Crosby NB, Dimitropoulou M, Georgoulis MK, Hergarten S, et al. 2016. 25 years of self-organized criticality: Solar and Astrophysics. *Space Sci Rev* **198**(1–4): 47–166. <https://doi.org/10.1007/s11214-014-0054-6>.
- Bak P. 1996. *How nature works: The science of self-organized criticality*. Springer, New York. ISBN 978-0387987385.
- Bak P, Tang C, Wiesenfeld K. 1987. Self-organized criticality: An explanation of the 1/f noise. *Phys Rev Lett* **59**(4): 381–384. <https://doi.org/10.1103/PhysRevLett.59.381>.
- Bak P, Tang C, Wiesenfeld K. 1988. Self-organized criticality. *Phys Rev A* **38**(1): 364–374. <https://doi.org/10.1103/PhysRevA.38.364>.
- Baker DN, Li X, Pulkkinen A, Ngwira CM, Mays ML, Galvin AB, Simunac KDC. 2013. A major solar eruptive event in July 2012: Defining extreme space weather scenarios. *Space Weather* **11**: 585–591. <https://doi.org/10.1002/swe.20097>.
- Barnes G. 2007. On the relationship between coronal magnetic null points and solar eruptive events. *Astrophys J Lett* **670**(1): L53–L56. <https://doi.org/10.1086/524107>.
- Barnes G, Leka KD. 2006. Photospheric magnetic field properties of flaring versus flare-quiet active regions. III. Magnetic charge topology models. *Astrophys J* **646**(2): 1303–1318. <https://doi.org/10.1086/504960>.
- Barnes G, Leka KD. 2008. Evaluating the performance of solar flare forecasting methods. *Astrophys J Lett* **688**: L107. <https://doi.org/10.1086/595550>.
- Barnes G, Leka KD, Schrijver CJ, Colak T, Qahwaji R, et al. 2016. A comparison of flare forecasting methods. I. Results from the “All-Clear” workshop. *Astrophys J* **829**: 89. <https://doi.org/10.3847/0004-637X/829/2/89>.
- Barnes G, Schanche N, Leka KD, Aggarwal A, Reeves K. 2017. A comparison of classifiers for solar energetic events. In: *Astroinformatics*, Brescia M, Djorgovski SG, Feigelson ED, Longo G, Cuvuoti S (Eds.), Cambridge University Press, Cambridge, Vol. **325**, pp. 201–204. <https://doi.org/10.1017/S1743921316012758>.
- Benvenuto F, Piana M, Campi C, Massone AM. 2018. A hybrid supervised/unsupervised machine learning approach to solar flare prediction. *Astrophys J* **853**(1): 90. <https://doi.org/10.3847/1538-4357/aaa23c>.
- Benz AO. 2008. Flare Observations. *Living Rev Sol Phys* **5**(1): 1. <https://doi.org/10.12942/lrsp-2008-1>.
- Berger MA. 2003. Topological quantities in magnetohydrodynamics. *Adv Nonlinear Dyn Ser Fluid Mech Astrophys Geophys* **20030424**: 345–374. <https://doi.org/10.1201/9780203493137.ch10>.
- Berger MA, Field GB. 1984. The topological properties of magnetic helicity. *J Fluid Mech* **147**: 133–148. <https://doi.org/10.1017/S0022112084002019>.
- Bezdek JC. 1981. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, Plenum Press, New York. ISBN 978-1-4757-0450-1.
- Bloomfield DS, Higgins PA, McAteer RTJ, Gallagher PT. 2012. Toward reliable benchmarking of solar flare forecasting methods. *Astrophys J Lett* **747**: L41. <https://doi.org/10.1088/2041-8205/747/2/L41>.
- Bobra MG, Couvidat S. 2015. Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophys J* **798**(2): 135. <https://doi.org/10.1088/0004-637X/798/2/135>.
- Bobra MG, Sun X, Hoeksema JT, Turmon M, Liu Y, Hayashi K, Barnes G, Leka KD. 2014. The helioseismic and magnetic imager (HMI) vector magnetic field pipeline: SHARPs – Space-Weather HMI Active Region Patches. *Sol Phys* **289**: 3549–3578. <https://doi.org/10.1007/s11207-014-0529-3>.
- Breiman L. 2001. Random forests. *Mach Learn* **45**(1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Broecker J. 2012. *Probability forecasts, chap. 7*, John Wiley & Sons, Oxford, pp. 119–139. ISBN: 9781119960003. <https://doi.org/10.1002/9781119960003.ch7>.
- Campi C, Benvenuto F, Massone AM, Bloomfield DS, Georgoulis MK, Piana M. 2019. Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence. *Astrophys J* **883**(2): 150. <https://doi.org/10.3847/1538-4357/ab3c26>.
- Camporeale E. 2019. The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather* **17**(8): 1166–1207. <https://doi.org/10.1029/2018SW002061>.
- Camporeale E, Wing S, Johnson J, Jackman CM, McGranaghan R. 2018. Space weather in the machine learning era: A multidisciplinary approach. *Space Weather* **16**: 2–4. <https://doi.org/10.1002/2017SW001775>.
- Carrington RC. 1859. Description of a singular appearance seen in the Sun on September 1, 1859. *Mon Not R Astron Soc* **20**: 13–15. <https://doi.org/10.1093/mnras/20.1.13>.
- Cliver EW, Dietrich WF. 2013. The 1859 space weather event revisited: Limits of extreme activity. *J Space Weather Space Clim* **3**(27): A31. <https://doi.org/10.1051/swsc/2013053>.
- Conlon PA, Gallagher PT, McAteer RTJ, Ireland J, Young CA, Kestener P, Hewett RJ, Maguire K. 2008. Multifractal properties of evolving active regions. *Sol Phys* **248**(2): 297–309. <https://doi.org/10.1007/s11207-007-9074-7>.
- Crosby NB, Aschwanden MJ, Dennis BR. 1993. Frequency distributions and correlations of solar X-ray flare parameters. *Sol Phys* **143**(2): 275–299. <https://doi.org/10.1007/BF00646488>.
- Dalmasse K, Aulanier G, Démoulin P, Kliem B, Török T, Pariat E. 2015. The origin of net electric currents in solar active regions. *Astrophys J* **810**(1): 17. <https://doi.org/10.1088/0004-637X/810/1/17>.
- De Vito E, Rosasco L, Caponnetto A, Piana M, Verri A. 2004. Some properties of regularized kernel methods. *J Mach Learn Res* **5** (Oct): 1363–1390.
- Deng N, Xu Y, Yang G, Cao W, Liu C, Rimmele TR, Wang H, Denker C. 2006. Multiwavelength study of flow fields in flaring super active region NOAA 10486. *Astrophys J* **644**(2): 1278. <https://doi.org/10.1086/503600>.

- Dietterich TG. 2000. Ensemble methods in machine learning. In: *International workshop on multiple classifier systems*, Kittler J, Roli F (Eds.), Springer-Verlag Berlin, Heidelberg, pp. 1–15.
- Drake JF. 1971. Characteristics of soft solar X-ray bursts. *Sol Phys* **16**: 152–185. <https://doi.org/10.1007/BF00154510>.
- Eastwood JP, Hapgood MA, Biffis E, Benedetti D, Bisi MM, Green L, Bentley RD, Burnett C. 2018. Quantifying the economic value of space weather forecasting for power grids: An exploratory study. *Space Weather* **16**: 2052–2067. <https://doi.org/10.1029/2018SW002003>.
- Efron B. 1982. *The Jackknife, the Bootstrap and other resampling plans*, Society for Industrial and Applied Mathematics, Philadelphia. <https://doi.org/10.1137/1.9781611970319>.
- Elman JL. 1990. Finding structure in time. *Cogn Sci* **14**(2): 179–211. https://doi.org/10.1207/s15516709cog1402_1.
- Elvidge S, Angling MJ. 2018. Using extreme value theory for determining the probability of Carrington-like solar flares. *Space Weather* **16**(4): 417–421. <https://doi.org/10.1002/2017SW001727>.
- Evgeniou T, Pontil M, Poggio T. 2000. Regularization networks and support vector machines. *Adv Comput Math* **13**(1): 1. <https://doi.org/10.1023/A:1018946025316>.
- Falconer DA, Moore RL, Gary GA. 2002. Correlation of the coronal mass ejection productivity of solar active regions with measures of their global nonpotentiality from vector magnetograms: Baseline results. *Astrophys J* **569**(2): 1016–1025. <https://doi.org/10.1086/339161>.
- Falconer DA, Moore RL, Barghouty AF, Khazanov I. 2012. Prior flaring as a complement to free magnetic energy for forecasting solar eruptions. *Astrophys J* **757**: 32. <https://doi.org/10.1088/0004-637X/757/1/32>.
- Fletcher L, Dennis BR, Hudson HS, Krucker S, Phillips K, et al. 2011. An observational overview of solar flares. *Space Sci Rev* **159**: 19–106. <https://doi.org/10.1007/s11214-010-9701-8>.
- Florios K, Kontogiannis I, Park S-H, Guerra JA, Benvenuto F, Bloomfield DS, Georgoulis MK. 2018. Forecasting solar flares using magnetogram-based predictors and machine learning. *Sol Phys* **293**(2): 28. <https://doi.org/10.1007/s11207-018-1250-4>.
- Freese S, Reddy AP, Lehnhardt K. 2016. Radiation impacts on human health during spaceflight beyond low earth orbit. *REACH* **2–4**: 1–7. <https://doi.org/10.1016/j.reach.2016.11.002>.
- Gary GA, Moore RL. 2004. Eruption of a multiple-turn helical magnetic flux tube in a large flare: Evidence for external and internal reconnection that fits the breakout model of solar magnetic eruptions. *Astrophys J* **611**(1): 545–556. <https://doi.org/10.1086/422132>.
- Georgoulis MK. 2005. Turbulence in the solar atmosphere: Manifestations and diagnostics via solar image processing. *Sol Phys* **228** (1–2): 5–27. <https://doi.org/10.1007/s11207-005-2513-4>.
- Georgoulis MK. 2008. Magnetic complexity in eruptive solar active regions and associated eruption parameters. *Geophys Res Lett* **35**(6): L06S02. <https://doi.org/10.1029/2007GL032040>.
- Georgoulis MK. 2011. Pre-eruption magnetic configurations in the active-region solar photosphere. In: *Physics of Sun and Star Spots*, Prasad Choudhary D, Strassmeier KG (Eds.), Vol. **273** of IAU Symposium, Cambridge University Press, Cambridge, pp. 495–498. <https://doi.org/10.1017/S1743921311015870>.
- Georgoulis MK. 2012. On our ability to predict major solar flares. *Astrophys Space Sci Proc* **30**: 93. https://doi.org/10.1007/978-3-642-29417-4_9.
- Georgoulis MK. 2013. Toward an efficient prediction of solar flares: Which parameters, and how? *Entropy* **15**(11): 5022–5052. <https://doi.org/10.3390/e15115022>.
- Georgoulis MK, Rust DM. 2007. Quantitative forecasting of major solar flares. *Astrophys J Lett* **661**: L109–L112. <https://doi.org/10.1086/518718>.
- Georgoulis MK, Titov VS, Mikić Z. 2012. Non-neutralized electric current patterns in solar active regions: Origin of the shear-generating Lorentz force. *Astrophys J* **761**: 61. <https://doi.org/10.1088/0004-637X/761/1/61>.
- Georgoulis MK, Nindos A, Zhang H. 2019. The source and engine of coronal mass ejections. *Philos Trans R Soc Lond Ser A* **377**(2148): 20180,094. <https://doi.org/10.1098/rsta.2018.0094>.
- Gibson SE, Fan Y. 2006. The partial expulsion of a magnetic flux rope. *Astrophys J Lett* **637**(1): L65–L68. <https://doi.org/10.1086/500452>.
- Gontikakis C, Kontogiannis I, Georgoulis MK, Guennou C, Syntelis P, Park SH, Buchlin E. 2020. Differential emission measure evolution as a precursor of solar flares. *arXiv e-prints*, arXiv:2011.06433. <https://arxiv.org/abs/2011.06433>.
- Goodfellow I, Bengio Y, Courville A. 2016. *Deep learning*, The MIT Press, Cambridge, MA, ISBN 0262035618.
- Green LM, Török T, Vršnak B, Manchester W, Veronig A. 2018. The origin, early evolution and predictability of solar eruptions. *Space Sci Rev* **214**(1): 46. <https://doi.org/10.1007/s11214-017-0462-5>.
- Guennou C, Auchère F, Soubrié E, Bocchialini K, Parenti S, Barbey N. 2012. On the accuracy of the differential emission measure diagnostics of solar plasmas. Application to SDO/AIA. I. Isothermal plasmas. *Astrophys J Suppl Ser* **203**: 25. <https://doi.org/10.1088/0067-0049/203/2/25>.
- Guennou C, Pariat E, Leake JE, Vilmer N. 2017. Testing predictors of eruptivity using parametric flux emergence simulations. *J Space Weather Space Clim* **7**: A17. <https://doi.org/10.1051/swsc/2017015>.
- Guerra JA, Pulkkinen A, Uritsky VM, Yashiro S. 2015. Spatio-temporal scaling of turbulent photospheric line-of-sight magnetic field in active region NOAA 11158. *Sol Phys* **290**: 335–350. <https://doi.org/10.1007/s11207-014-0636-1>.
- Guerra JA, Park SH, Gallagher PT, Kontogiannis I, Georgoulis MK, Bloomfield DS. 2018. Active region photospheric magnetic properties derived from line-of-sight and radial fields. *Sol Phys* **293**(1): 9. <https://doi.org/10.1007/s11207-017-1231-z>.
- Haggerty DK, Roelof EC. 2002. Impulsive near-relativistic solar electron events: Delayed injection with respect to solar electromagnetic emission. *Astrophys J* **579**: 841–853. <https://doi.org/10.1086/342870>.
- Hagyard MJ, Smith JB, Teuber D, West EA. 1984. A quantitative study relating observed shear in photospheric magnetic fields to repeated flaring. *Sol Phys* **91**(1): 115–126. <https://doi.org/10.1007/BF00213618>.
- Haynes AL, Parnell CE. 2007. A trilinear method for finding null points in a three-dimensional vector space. *Phys Plasma* **14**(8): 082,107. <https://doi.org/10.1063/1.2756751>.
- Hewett RJ, Gallagher PT, McAteer RTJ, Young CA, Ireland J, Conlon PA, Maguire K. 2008. Multiscale analysis of active region evolution. *Sol Phys* **248**: 311–322. <https://doi.org/10.1007/s11207-007-9028-0>.
- Hock RA. 2012. The role of solar flares in the variability of the extreme ultraviolet solar spectral irradiance. *Ph.D. thesis*, University of Colorado, Boulder.
- Hoeksema JT, Liu Y, Hayashi K, Sun X, Schou J, et al. 2014. The helioseismic and magnetic imager (HMI) vector magnetic field pipeline: Overview and performance. *Sol Phys* **289**(9): 3483–3530. <https://doi.org/10.1007/s11207-014-0516-8>.
- Howard R, Severny A. 1963. Solar magnetic fields and the great flare of July 16, 1959. *Astrophys J* **137**: 1242. <https://doi.org/10.1086/147600>.

- Jolliffe IT, Stephenson DB (Eds). 2012. *Forecast verification: A practitioner's guide in atmospheric science*, 2nd edn. John Wiley & Sons, Oxford. <https://doi.org/10.1002/9781119960003>.
- Jordan MI. 1997. Serial order: A parallel distributed processing approach. In: *Advances in psychology*, Donahoe JW, Packard Dorsel V (Eds.), Vol. 121, Elsevier, Amsterdam, pp. 471–495. [https://doi.org/10.1016/S0166-4115\(97\)80111-2](https://doi.org/10.1016/S0166-4115(97)80111-2).
- Kadanoff LP, Nagel SR, Wu L, Zhou S-M. 1989. Scaling and universality in avalanches. *Phys Rev A* 39: 6524–6537. <https://doi.org/10.1103/PhysRevA.39.6524>.
- Kontogiannis I, Georgoulis MK, Park S-H, Guerra JA. 2017. Non-neutralized electric currents in solar active regions and flare productivity. *Sol Phys* 292: 159. <https://doi.org/10.1007/s11207-017-1185-1>.
- Kontogiannis I, Georgoulis MK, Park S-H, Guerra JA. 2018. Testing and improving a set of morphological predictors of flaring activity. *Sol Phys* 293: 96. <https://doi.org/10.1007/s11207-018-1317-2>.
- Kontogiannis I, Georgoulis MK, Guerra JA, Park S-H, Bloomfield DS. 2019. Which photospheric characteristics are most relevant to active-region coronal mass ejections? *Sol Phys* 294(9): 130. <https://doi.org/10.1007/s11207-019-1523-6>.
- Korsós MB, Erdélyi R. 2016. On the state of a solar active region before flares and CMEs. *Astrophys J* 823: 153. <https://doi.org/10.3847/0004-637X/823/2/153>.
- Korsós MB, Baranyi T, Ludmány A. 2014. Pre-flare dynamics of sunspot groups. *Astrophys J* 789: 107. <https://doi.org/10.1088/0004-637X/789/2/107>.
- Korsós MB, Ludmány A, Erdélyi R, Baranyi T. 2015. On flare predictability based on sunspot group evolution. *Astrophys J Lett* 802: L21. <https://doi.org/10.1088/2041-8205/802/2/L21>.
- Krall KR, Smith JB, Hagyard MJ, West EA, Cummings NP. 1982. Vector magnetic field evolution, energy storage, and associated photospheric velocity shear within a flare-productive active region. *Sol Phys* 79(1): 59–75. <https://doi.org/10.1007/BF00146973>.
- Krishnapuram R, Keller JM. 1996. The Possibilistic C-Means Algorithm: Insights and Recommendations. *Trans Fuz Syst* 4(3): 385–393. <https://doi.org/10.1109/91.531779>.
- Kusano K, Maeshiro T, Yokoyama T, Sakurai T. 2002. Measurement of magnetic helicity injection and free energy loading into the solar corona. *Astrophys J* 577(1): 501. <https://doi.org/10.1086/342171>.
- Kusano K, Bamba Y, Yamamoto TT, Iida Y, Toriumi S, Asai A. 2012. Magnetic field structures triggering solar flares and coronal mass ejections. *Astrophys J* 760(1): 31. <https://doi.org/10.1088/0004-637X/760/1/31>.
- Leake JE, Linton MG, Török T. 2013. Simulations of emerging magnetic flux. I. The formation of stable coronal flux ropes. *Astrophys J* 778(2): 99. <https://doi.org/10.1088/0004-637X/778/2/99>.
- Leake JE, Linton MG, Antiochos SK. 2014. Simulations of emerging magnetic flux. II. The formation of unstable coronal flux ropes and the initiation of coronal mass ejections. *Astrophys J* 787(1): 46. <https://doi.org/10.1088/0004-637X/787/1/46>.
- Lee K, Moon YJ, Lee J-Y, Lee K-S, Na H. 2012. Solar flare occurrence rate and probability in terms of the sunspot classification supplemented with sunspot area and its changes. *Sol Phys* 281(2): 639–650. <https://doi.org/10.1007/s11207-012-0091-9>.
- Leka KD, Barnes G. 2003a. Photospheric magnetic field properties of flaring versus flare-quiet active regions. I. Data, general approach, and sample results. *Astrophys J* 595(2): 1277–1295. <https://doi.org/10.1086/377511>.
- Leka KD, Barnes G. 2003b. Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant analysis. *Astrophys J* 595: 1296–1306. <https://doi.org/10.1086/377512>.
- Leka KD, Barnes G. 2007. Photospheric magnetic field properties of flaring versus flare-quiet active regions. IV. A statistically significant sample. *Astrophys J* 656(2): 1173–1186. <https://doi.org/10.1086/510282>.
- Leka KD, Barnes G, Wagner E. 2018. The NWRA classification infrastructure: Description and extension to the discriminant analysis flare forecasting system (DAFFS). *Journal of Space Weather and Space Climate* 8: A25. <https://doi.org/10.1051/swsc/2018004>.
- Leka KD, Park S-H, Kusano K, Andries J, Barnes G, et al. 2019a. A comparison of flare forecasting methods. II. Benchmarks, metrics, and performance results for operational solar flare forecasting systems. *Astrophys J Suppl Ser* 243(2): 36. <https://doi.org/10.3847/1538-4365/ab2e12>.
- Leka KD, Park S-H, Kusano K, Andries J, Barnes G, et al. 2019b. A comparison of flare forecasting methods. III. Systematic behaviors of operational solar flare forecasting systems. *Astrophys J* 881(2): 101. <https://doi.org/10.3847/1538-4357/ab2e11>.
- Li R, Cui Y, He H, Wang H. 2008. Application of support vector machine combined with K-nearest neighbors in solar flare and solar proton events forecasting. *Adv Space Res* 42(9): 1469–1474. <https://doi.org/10.1016/j.asr.2007.12.015>.
- Li X, Zheng Y, Wang X, Wang L. 2020. Predicting solar flares using a novel deep convolutional neural network. *Astrophys J* 891(1): 10. <https://doi.org/10.3847/1538-4357/ab6d04>.
- Liaw A, Wiener M. 2002. Classification and regression by randomForest. In: *The Newsletter of the R Project Volume*, The R-Foundation CRAN R-Project, 2 p. <https://cran.r-project.org>.
- Linan L, Pariat E, Moraitis K, Valori G, Leake J. 2018. Time variations of the nonpotential and volume-threading magnetic helicities. *Astrophys J* 865(1): 52. <https://doi.org/10.3847/1538-4357/aadae7>.
- Liu Y. 2008. Magnetic field overlying solar eruption regions and kink and torus instabilities. *Astrophys J Lett* 679(2): L151. <https://doi.org/10.1086/589282>.
- Liu L, Wang Y, Wang J, Shen C, Ye P, Liu R, Chen J, Zhang Q, Wang S. 2016. Why is a flare-rich active region CME-poor? *Astrophys J* 826(2): 119. <https://doi.org/10.3847/0004-637X/826/2/119>.
- Longadge R, Dongre S. 2013. Class imbalance problem in data mining review. *Int J Comput Sci Netw (IJCSN)* 2(1). ISSN 2277-5420. arXiv e-print article: <https://arxiv.org/abs/1305.1707>.
- Lu ET, Hamilton RJ. 1991. Avalanches and the distribution of solar flares. *Astrophys J Lett* 380: L89. <https://doi.org/10.1086/186180>.
- Lu ET, Hamilton RJ, McTiernan JM, Bromund KR. 1993. Solar flares and avalanches in driven dissipative systems. *Astrophys J* 412: 841. <https://doi.org/10.1086/172966>.
- MacAlester MH, Murtagh W. 2014. Extreme space weather impact: An emergency management perspective. *Space Weather* 12(8): 530–537. <https://doi.org/10.1002/2014SW001095>.
- Malandraki OE, Crosby NB. 2018. Solar particle radiation storms forecasting and analysis. In: Vol. 444 of *Astrophysics and Space Science Library*, Springer International Publishing, Switzerland. <https://doi.org/10.1007/978-3-319-60051-2>.
- Martens PCH, Attrill GDR, Davey AR, Engell A, Farid S, et al. 2012. Computer vision for the solar dynamics observatory (SDO). *Sol Phys* 275(1–2): 79–113. <https://doi.org/10.1007/s11207-010-9697-y>.
- Mason JP, Hoeksema JT. 2010. Testing automated solar flare forecasting with 13 years of michelson doppler imager magnetograms. *Astrophys J* 723: 634–640. <https://doi.org/10.1088/0004-637X/723/1/634>.
- Massone A, Studer L, Masulli F. 2006. Possibilistic clustering approach to trackless ring pattern recognition in RICH counters. *Int J Approx Reason* 41(2): 96–109. Advances in Fuzzy Sets and Rough Sets. <http://doi.org/10.1016/j.ijar.2005.06.012>.

- Massone AM, Piana M, the FLARECAST Consortium. 2018. Chapter 14 – machine learning for flare forecasting. In: *Machine learning techniques for space weather*, Camporeale E, Wing S, Johnson JR (Eds.), Elsevier Amsterdam, Oxford, Cambridge MA, pp. 355–364. ISBN 978-0-12-811788-0. <https://doi.org/10.1016/B978-0-12-811788-0.00014-7>.
- Mays ML, Taktakishvili A, Pulkkinen A, MacNeice PJ, Rastätter L, et al. 2015. Ensemble modeling of CMEs using the WSA-ENLIL + Cone model. *Sol Phys* **290**: 1775–1814. <https://doi.org/10.1007/s11207-015-0692-1>.
- McAteer RTJ, Aschwanden MJ, Dimitropoulou M, Georgoulis MK, Pruessner G, Morales L, Ireland J, Abramenko V. 2016. 25 years of self-organized criticality: Numerical detection methods. *Space Sci Rev* **198**(1–4): 217–266. <https://doi.org/10.1007/s11214-015-0158-7>.
- McCloskey AE, Gallagher PT, Bloomfield DS. 2016. Flaring rates and the evolution of sunspot group McIntosh classifications. *Sol Phys* **291**(6): 1711–1738. <https://doi.org/10.1007/s11207-016-0933-y>.
- McIntosh PS. 1990. The classification of sunspot groups. *Sol Phys* **125**: 251–267. <https://doi.org/10.1007/BF00158405>.
- Merceret FJ, O'Brien TP, Roeder WP, Huddleston LL, Bauman WH, Jedlovec GJ. 2013. Transitioning research to operations: Transforming the “Valley of Death” into a “Valley of Opportunity”. *Space Weather* **11**(11): 637–640. <https://doi.org/10.1002/swe.20099>.
- Meyers RA (ed). 2011. *Extreme environmental events*, Springer, New York. <https://doi.org/10.1007/978-1-4419-7695-6>.
- Moraitis K, Tziotziou K, Georgoulis MK, Archontis V. 2014. Validation and benchmarking of a practical free magnetic energy and relative magnetic helicity budget calculation in solar magnetic structures. *Sol Phys* **122**: 4453–4480. <https://doi.org/10.1007/s11207-014-0590-y>.
- Möstl C, Amla K, Hall JR, Liewer PC, De Jong EM, et al. 2014. Connecting speeds, directions and arrival times of 22 coronal mass ejections from the sun to 1 AU. *Astrophys J* **787**: 119. <https://doi.org/10.1088/0004-637X/787/2/119>.
- Muranushi T, Shibayama T, Muranushi YH, Isobe H, Nemoto S, Komazaki K, Shibata K. 2015. UFCORIN: A fully automated predictor of solar flares in GOES X-ray flux. *Space Weather* **13**(11): 778–796. <https://doi.org/10.1002/2015sw001257>.
- Murray SA, Bingham S, Sharpe M, Jackson DR. 2017. Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather* **15**: 577–588. <https://doi.org/10.1002/2016SW001579>.
- Murray SA, Guerra JA, Zucca P, Park S-H, Carley EP, Gallagher PT, Vilmer N, Bothmer V. 2018. Connecting coronal mass ejections to their solar active region sources: Combining results from the HELCATS and FLARECAST projects. *Sol Phys* **293**: 60. <https://doi.org/10.1007/s11207-018-1287-4>.
- Nishizuka N, Sugiura K, Kubo Y, Den M, Watari S, Ishii M. 2017. Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms. *Astrophys J* **835**(2): 156. <https://doi.org/10.3847/1538-4357/835/2/156>.
- Nishizuka N, Sugiura K, Kubo Y, Den M, Ishii M. 2018. Deep flare net (DeFN) model for solar flare prediction. *Astrophys J* **858**(2): 113. <https://doi.org/10.3847/1538-4357/aab9a7>.
- NRC. 2008. *Severe space weather events: Understanding societal and economic impacts: A workshop report*. The National Academies Press, Washington, DC. ISBN: 978-0-309-12769-1. <https://doi.org/10.17226/12507>.
- Opgenoorth HJ, Wimmer-Schweingruber RF, Belehaki A, Berghmans D, Hapgood M, et al. 2019. Assessment and recommendations for a consolidated European approach to space weather – as part of a global space weather effort. *J Space Weather Space Clim* **9**: A37. <https://doi.org/10.1051/swsc/2019033>.
- Oughton EJ, Skelton A, Horne RB, Thomson AWP, Gaunt CT. 2017. Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure. *Space Weather* **15**: 65–83. <https://doi.org/10.1002/2016SW001491>.
- Pariat E, Leake JE, Valori G, Linton MG, Zuccarello FP, Dalmasse K. 2017. Relative magnetic helicity as a diagnostic of solar eruptivity. *A&A* **601**: A125. <https://doi.org/10.1051/0004-6361/201630043>.
- Park S-H, Chae J, Wang H. 2010. Productivity of solar flares and magnetic helicity injection in active regions. *Astrophys J* **718**(1): 43. <https://doi.org/10.1088/0004-637X/718/1/43>.
- Park S-H, Cho K-S, Bong S-C, Kumar P, Chae J, Liu R, Wang H. 2012. The occurrence and speed of CMEs related to two characteristic evolution patterns of helicity injection in their solar source regions. *Astrophys J* **750**: 48. <https://doi.org/10.1088/0004-637X/750/1/48>.
- Park S-H, Guerra JA, Gallagher PT, Georgoulis MK, Bloomfield DS. 2018. Photospheric shear flows in solar active regions and their relation to flare occurrence. *Sol Phys* **293**: 114. <https://doi.org/10.1007/s11207-018-1336-z>.
- Park S-H, Leka KD, Kusano K, Andries J, Barnes G, et al. 2020. A comparison of flare forecasting methods. IV. Evaluating consecutive-day forecasting patterns. *Astrophys J* **890**(2): 124. <https://doi.org/10.3847/1538-4357/ab65f0>.
- Patsourakos S, Vourlidas A, Török T, Kliem B, Antiochos SK, et al. 2020. Decoding the pre-eruptive magnetic field configurations of coronal mass ejections. *Space Sci Rev* **216**(8): 131. <https://doi.org/10.1007/s11214-020-00757-9>.
- Pesnell WD, Thompson BJ, Chamberlin PC. 2012. The solar dynamics observatory (SDO). *Sol Phys* **275**(1–2): 3–15. <https://doi.org/10.1007/s11207-011-9841-3>.
- Piana M, Campi C, Benvenuto F, Guastavino S, Massone AM. 2019. Flare forecasting and feature ranking using SDO/HMI data. *Nuovo Cimento C* **42**(30): 4. <https://doi.org/10.1393/ncc/i2019-19030-1>.
- Pontin DI, Priest ER, Galsgaard K. 2013. On the nature of reconnection at a solar coronal null point above a separatrix dome. *Astrophys J* **774**(2): 154. <https://doi.org/10.1088/0004-637X/774/2/154>.
- Qahwaji R, Colak T. 2007. Automatic short-term solar flare prediction using machine learning and sunspot associations. *Sol Phys* **241**(1): 195–211. <https://doi.org/10.1007/s11207-006-0272-5>.
- Qahwaji R, Colak T, Al-Omari M, Ipson S. 2008. Automated prediction of CMEs using machine learning of CME – flare associations. *Sol Phys* **248**: 471–483. <https://doi.org/10.1007/s11207-007-9108-1>.
- Reames DV (ed). 2017. Solar energetic particles. In: Vol. **932** of *Lecture Notes in Physics*, Berlin Springer Verlag. <https://doi.org/10.1007/978-3-319-50871-9>.
- Richardson DS. 2012. *Economic value and skill*, Chap. **9**, John Wiley & Sons, Oxford, pp. 167–184. ISBN 9781119960003. <https://doi.org/10.1002/9781119960003.ch9>.
- Riley P, Love JJ. 2017. Extreme geomagnetic storms: Probabilistic forecasts and their uncertainties. *Space Weather* **15**: 53–64. <https://doi.org/10.1002/2016SW001470>.
- Rosasco L, Vito ED, Caponnetto A, Piana M, Verri A. 2004. Are loss functions all the same? *Neural Comput* **16**(5): 1063–1076. <https://doi.org/10.1162/089976604773135104>.
- Rosner R, Vaiana GS. 1978. Cosmic flare transients: Constraints upon models for energy storage and release derived from the event frequency distribution. *Astrophys J* **222**: 1104–1108. <https://doi.org/10.1086/156227>.

- Rudin C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Mach Intel* **1**: 206–215. <https://doi.org/10.1038/s42256-019-0048-x>.
- Rust DM, Haggerty DK, Georgoulis MK, Sheeley NR, Wang Y-M, DeRosa ML, Schrijver CJ. 2008. On the solar origins of open magnetic fields in the heliosphere. *Astrophys J* **687**: 635–645. <https://doi.org/10.1086/592017>.
- Scherrer PH, Schou J, Bush RI, Kosovichev AG, Bogart RS, et al. 2012. The helioseismic and magnetic imager (HMI) Investigation for the solar dynamics observatory (SDO). *Sol Phys* **275**: 207–227. <https://doi.org/10.1007/s11207-011-9834-2>.
- Scholkopf B, Smola AJ. 2001. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*, The MIT Press, Cambridge, MA. ISBN 978-0262536578.
- Schou J, Scherrer PH, Bush RI, Wachter R, Couvidat S, et al. 2012. Design and ground calibration of the helioseismic and magnetic imager (HMI) instrument on the solar dynamics observatory (SDO). *Sol Phys* **275(1–2)**: 229–259. <https://doi.org/10.1007/s11207-011-9842-2>.
- Schrijver CJ. 2007. A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *ApJL* **655**: L117–L120. <https://doi.org/10.1086/511857>.
- Schrijver CJ. 2009. Driving major solar flares and eruptions: A review. *Adv Space Res* **43(5)**: 739–755. <https://doi.org/10.1016/j.asr.2008.11.004>.
- Schrijver CJ, De Rosa ML, Title AM, Metcalf TR. 2005. The nonpotentiality of active-region coronae and the dynamics of the photospheric magnetic field. *Sol Phys* **628(1)**: 501–513. <https://doi.org/10.1086/430733>.
- Schrijver CJ, Kauristie K, Aylward AD, Denardini CM, Gibson SE, et al. 2015. Understanding space weather to shield society: A global road map for 2015–2025 commissioned by COSPAR and ILWS. *Adv Space Res* **55(12)**: 2745–2807. <https://doi.org/10.1016/j.asr.2015.03.023>.
- Sharma AS, Bunde A, Dimri VP, Baker DN (Eds). 2012. *Extreme events and natural hazards: The complexity perspective*. American Geophysical Union, Washington DC. <https://doi.org/10.1029/gm196>.
- Sharpe MA, Murray SA. 2017. Verification of space weather forecasts issued by the Met Office Space Weather Operations Centre. *Space Weather* **15(10)**: 1383–1395. <https://doi.org/10.1002/2017SW001683>.
- Shu K, Cui L, Wang S, Lee D, Liu H. 2019. *DEFEND: explainable fake news detection*. Association for Computing Machinery, New York, NY, USA. ISBN 9781450362016. <https://doi.org/10.1145/3292500.3330935>.
- Song H, Tan C, Jing J, Wang H, Yurchyshyn V, Abramenko V. 2009. Statistical assessment of photospheric magnetic features in imminent solar flare predictions. *Sol Phys* **254(1)**: 101–125. <https://doi.org/10.1007/s11207-008-9288-3>.
- Strugarek A, Charbonneau P. 2014. Predictive capabilities of avalanche models for solar flares. *Sol Phys* **289(11)**: 4137–4150. <https://doi.org/10.1007/s11207-014-0570-2>.
- Syntelis P, Gontikakis C, Patsourakos S, Tsinganos K. 2016. The spectroscopic imprint of the pre-eruptive configuration resulting into two major coronal mass ejections. *A&A* **588**: A16. <https://doi.org/10.1051/0004-6361/201526829>.
- Thompson WT. 2006. Coordinate systems for solar image data. *A&A* **449(2)**: 791–803. <https://doi.org/10.1051/0004-6361:20054262>.
- Toriumi S, Wang H. 2019. Flare-productive active regions. *Living Rev Sol Phys* **16(1)**: 3. <https://doi.org/10.1007/s41116-019-0019-7>.
- Török T, Leake JE, Titov VS, Archontis V, Mikić Z, Linton MG, Dalmasse K, Aulanier G, Kliem B. 2014. Distribution of electric currents in solar active regions. *Astrophys J Lett* **782(1)**: L10. <https://doi.org/10.1088/2041-8205/782/1/L10>.
- Ugarte-Urra I, Warren HP, Winebarger AR. 2007. The magnetic topology of coronal mass ejection sources. *Astrophys J* **662(2)**: 1293–1301. <https://doi.org/10.1086/514814>.
- van Ballegooijen AA, Martens PCH. 1989. Formation and eruption of solar prominences. *Astrophys J* **343**: 971. <https://doi.org/10.1086/167766>.
- Verbeke C, Mays ML, Temmer M, Bingham S, Steenburgh R, et al. 2019. Benchmarking CME arrival time and impact: Progress on metadata, metrics, and events. *Space Weather* **17(1)**: 6–26. <https://doi.org/10.1029/2018SW002046>.
- Vlahos L, Georgoulis M, Kluiving R, Paschos P. 1995. The statistical flare. *A&A* **299**: 897.
- Wang S, Liu C, Deng N, Wang H. 2014. Sudden photospheric motion and sunspot rotation associated with the X2.2 Flare on 2011 February 15. *Astrophys J Lett* **782(2)**: L31. <https://doi.org/10.1088/2041-8205/782/2/L31>.
- Welsch BT, Li Y, Schuck PW, Fisher GH. 2009. What is the relationship between photospheric flow fields and solar flares? *Astrophys J* **705(1)**: 821–843. <https://doi.org/10.1088/0004-637X/705/1/821>.
- Woodcock F. 1976. The evaluation of yes/no forecasts for scientific and administrative purposes. *Month Weather Rev* **104(10)**: 1209–1214. [https://doi.org/10.1175/1520-0493\(1976\)104<1209:TEOYFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<1209:TEOYFF>2.0.CO;2).
- Wu TT, Chen YF, Hastie T, Sobel E, Lange K. 2009. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* **25(6)**: 714–721. <https://doi.org/10.1093/bioinformatics/btp041>.
- Yan XL, Wang JC, Pan GM, Kong DF, Xue ZK, Yang LH, Li QL, Feng XS. 2018. Successive X-class flares and coronal mass ejections driven by shearing motion and sunspot rotation in active region NOAA 12673. *Astrophys J* **856(1)**: 79. <https://doi.org/10.3847/1538-4357/aab153>.
- Yashiro S, Gopalswamy N, Akiyama S, Michalek G, Howard RA. 2005. Visibility of coronal mass ejections as a function of flare location and intensity. *J Geophys Res (Space Phys)* **110(A12)**: A12S05. <https://doi.org/10.1029/2005JA011151>.
- Yu D, Huang X, Wang H, Cui Y. 2009. Short-term solar flare prediction using a sequential supervised learning method. *Sol Phys* **255(1)**: 91–105. <https://doi.org/10.1007/s11207-009-9318-9>.
- Yuan M, Lin Y. 2006. Model selection and estimation in regression with grouped variables. *J Roy Stat Soc: Ser B (Stat Methodol)* **68(1)**: 49–67. <https://doi.org/10.1111/j.1467-9868.2005.00532.x>.
- Zirin H, Liggett MA. 1987. Delta spots and great flares. *Sol Phys* **113(1–2)**: 267–283. <https://doi.org/10.1007/BF00147707>.
- Zirin H, Marquette W. 1991. BEARALERTS: A successful flare prediction system. *Sol Phys* **131(1)**: 149–164. <https://doi.org/10.1007/bf00151751>.
- Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *J Roy Stat Soc: Ser B (Stat Methodol)* **67(2)**: 301–320. <https://doi.org/10.1111/j.1467-7412.05.67301>.
- Zuccarello FP, Seaton DB, Mierla M, Poedts S, Rachmeler LA, Romano P, Zuccarello F. 2014. Observational evidence of torus instability as trigger mechanism for coronal mass ejections: The 2011 August 4 filament eruption. *Astrophys J* **785(2)**: 88. <https://doi.org/10.1088/0004-637X/785/2/88>.
- Zuccarello FP, Pariat E, Valori G, Linan L. 2018. Threshold of non-potential magnetic helicity ratios at the onset of solar eruptions. *Astrophys J* **863(1)**: 41. <https://doi.org/10.3847/1538-4357/aacdfc>.

Appendix A

Open access of FLARECAST data, codes and infrastructure

A.1 Data

Access to the FLARECAST property database is provided via <https://api.flarecast.eu/property/ui/>, while the prediction database is accessible via <https://api.flarecast.eu/prediction/ui/>. Both services are Swagger-based APIs, enabling data access in both human and machine readable format. They provide a collection of API routes to retrieve and analyse the FLARECAST data. While analysis is useful for developers and system administrators, the *View* routes are intended for end users. The graphical interfaces at these URLs serve two purposes. First, one can retrieve and verify data manually. Second, the interfaces print out a URL than can be used as query template and can also be integrated in any programming language supporting HTTP file access.

The most useful *View* routes of the property database service are:

- *data set/list*, that provides the list of all registered data sets. FLARECAST primarily works with the *production_03* (1 September 2012 to 12 April 2016 and 9 September 2017 to 30 January 2019) and *questionable_02* (13 April 2016 to 8 September 2017) data sets. The second data set includes HMI NRT data that may be problematic, according to a JSOC release²¹.
- */property/types*, that provides the list of individual property names. This is useful when one is interested in the acquisition of certain properties only.
- */data set/{data set}/list*, that provides the main access to the property database. Various options can be used to refine the search, while blank data requests are also possible. As an example, a general API URL for all property values within the 2-day interval of 1 June 2014 to 23 June 2014 at all available cadences reads: [https://api.flarecast.eu/property/region/production_03/list?cadence=all&exclude_higher_cadences=false&time_start=between\(2014-06-01,2014-06-03\)&property_type=*®ion_fields=*](https://api.flarecast.eu/property/region/production_03/list?cadence=all&exclude_higher_cadences=false&time_start=between(2014-06-01,2014-06-03)&property_type=*®ion_fields=*).

One may refine the search as needed, in terms of *cadence*, time interval (via *time_start*), or specific property (via *property_type*).

The most useful *View* route of the prediction database service is:

- */prediction/list_v2*, that provides the main access to the prediction database. As with the property database, various options can be used to refine the search. As an example, an API URL for all available predictions within the 5-month interval of 1 January 2018 to 1 June 2018 reads: [https://api.flarecast.eu/prediction/prediction/list_v2?include_flare_associations=true&algorithm_config_version=latest&prediction_time_start=between\(2018-01-01,2018-06-01\)](https://api.flarecast.eu/prediction/prediction/list_v2?include_flare_associations=true&algorithm_config_version=latest&prediction_time_start=between(2018-01-01,2018-06-01)).

²¹ <http://jsoc.stanford.edu/jsocwiki/ModLRcalibration>

The search window can be adapted by *prediction_time_start*. Each entry provides the algorithm configuration name (*algorithm_config_name*) and id (*algorithm_config_id*), along with all other applicable information. One can look into certain algorithm configuration names to refine the search further by setting the *algorithm_config_name* field accordingly and obtain the respective API route.

Most other routes in the property and prediction database services are still used by some of the user interfaces and some debugging scripts, but are not intended for end users.

A.2 Codes

All FLARECAST source codes and computational infrastructure are public and can be found in the project's Bitbucket repository: <https://dev.flarecast.eu/stash/projects/>. The repository includes different directories, namely:

- **Download:** algorithms for downloading SWPC and flare staging data.
- **Property Extraction:** codes and algorithms for the extraction of all properties included in [Table 1](#).
- **Flare Prediction:** source codes for the machine learning algorithms included in [Table 2](#).
- **FLARECAST Datamodel:** codes for defining and implementing the project's data model ([Sect. 2.4.8](#)).
- **FLARECAST Infrastructure:** codes for the project's computational infrastructure ([Sect. 3](#)).

Algorithms typically include brief usage notes, examples and information on each applicable programming language.

A.3 Infrastructure installation

The first dependency is the installation of a Docker engine from <https://www.docker.com>. Docker is an open source software that offers an ecosystem in which different Docker containers co-exist and function independently. Given that each FLARECAST Docker container contains a different routine or algorithm written in different programming languages (IDL, Python, C, or R), the second dependency is that the installation server must be equipped with licenses of these languages to be able to edit / modify the infrastructure. With Docker engine installation in place, the FLARECAST infrastructure is installed in a Unix/Linux command line environment as follows:

- Download the installer script:
curl -o infrastructure.sh <https://dev.flarecast.eu/stash/projects/INFRA/repos/dev-infra/browse/infrastructure.sh?raw>
- Assign execution mode to the script:
chmod a+x infrastructure.sh
- Setup the infrastructure or update to the most recent version:
./infrastructure.sh update
- Finally, run the development infrastructure:
./infrastructure.sh run

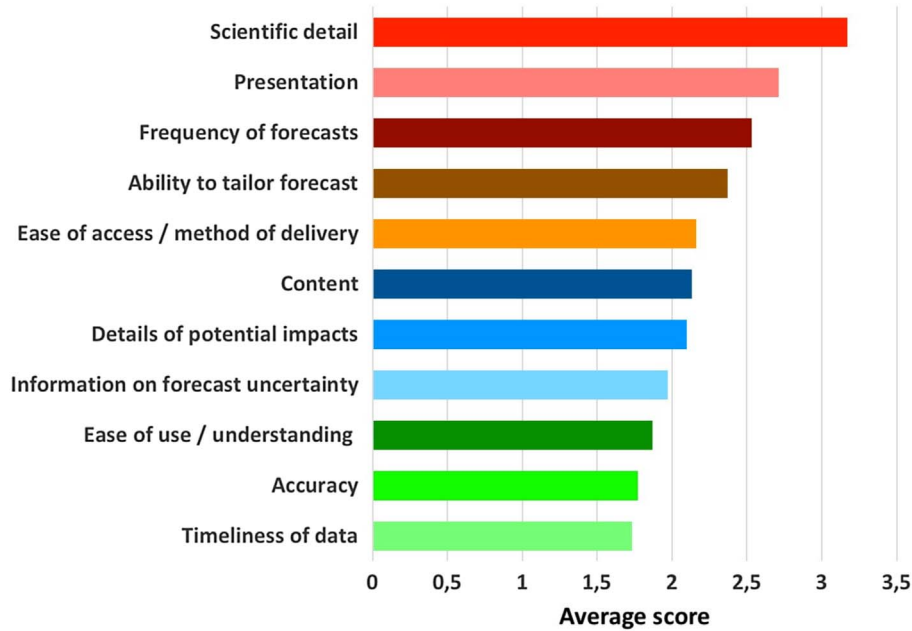


Fig. B.1. Summary of responses to the question, “Which factors are/would be important to you in a flare forecasting service?” Lowest average score corresponds to the most important factor.

A.4 Licensing

EU’s OpenAIRE policy requires that all source code and data models must be openly accessible to any interested person or entity, both within the EU and worldwide. Therefore, the software and data licenses chosen by the FLARECAST Consortium aim to be as permissive as possible, at the same time being free of any liability and provided “as is”. The majority of the written code is available under a BSD-2 (<https://opensource.org/licenses/BSD-2-Clause>) license. Some of the algorithms, however, depend on third-party libraries that use a less permissive GPL licence (<https://www.gnu.org/licenses/gpl-3.0.de.html>) and thus these algorithms need to be released under GPL, as well. All FLARECAST data products are provided under the fully open ODC PDDL (<https://www.opendatacommons.org/licenses/pddl/1-0/index.html>) license.

Appendix B

The FLARECAST user survey

The FLARECAST user survey was carried out in Autumn 2016, with around 100 users approached and 31 responding. Around 60% indicated that they were using flare forecast or alert services. Most thought these services were useful but around a quarter did not know whether their preferred service was accurate or inaccurate, which points to a need for better informed end users. Asked whether they would recommend such a service, around two thirds gave a “passive” response, indicating that they neither strongly recommend, nor strongly not recommend, flare forecasting services.

Users were asked if they planned to use flare forecasting services differently in the future. Five of them provided further details of how they would do this. While the assumption is that most users are in the aviation or defence industries, it is interesting that out of these five comments, one was from a satellite user and two from GNSS users. Suggested approaches included:

- Developing an operational response to space weather events, and a change in practices to take advantage of increased precision and advance warnings.
- Comparison with other flare prediction tools, such as ASSA²².
- Examining correlations between flares and ionospheric TEC changes.
- Checking the impact of space weather disturbances on GNSS systems.
- Allowing the spacecraft operator to tailor services for spacecraft control aspects.

The users were then asked what factors would be most important to them in a flare forecasting service. Timeliness and accuracy were found to be most important to a user, while scientific details were deemed least important. A summary of all responses appears in [Figure B.1](#).

Most flare forecasts focus on the occurrence of flares defined by the NOAA R scales (e.g., M-class, X-class), but 77% of responders said forecasts of “All Clear” periods are also important. A similar number knew about the NOAA R scales, but 12.5% said that they were not useful for their purposes. The following reasons were given:

²² <http://spaceweather.rra.go.kr/models/assa>

- The R scales classify impacts to radio systems, so are of some interest to us (a Telecoms company), but we are also interested in space weather impacts to other systems such as power distribution.
- The scale is useful globally, but too coarse at local (i.e., country) level.
- Currently we are not reacting to NOAA scales. A decision was taken to react to MOSWOC sourced indices.

Finally, the survey came up with suggestions of other useful points for discussion during the First Stakeholders' Workshop. Suggestions included:

- Details on verification of results and methods.
- Explanations for non-scientists; international coordination to raise awareness.
- Coupling to D-RAP²³; specific solar radio burst forecasts/alerts.
- The state of the science and roadblocks for forecast developments.
- Better understanding of impact / consequence of space weather events, and presentation of this for non-experts.
- Clear presentation of applicable uncertainties.

Appendix C

FLARECAST-related refereed publications

A total of eighteen (18) refereed publications, including this overview, were fully or partially supported by the FLARECAST project. The other seventeen (17) are (alphabetically by first author): [Barnes et al. \(2016\)](#), [Benvenuto et al. \(2018\)](#), [Campi et al. \(2019\)](#), [Florios et al. \(2018\)](#), [Gontikakis et al. \(2020\)](#), [Guennou et al. \(2017\)](#), [Guerra et al. \(2018\)](#), [Kontogiannis et al. \(2017, 2018, 2019\)](#), [Massone et al. \(2018\)](#), [McCloskey et al. \(2016\)](#), [Murray et al. \(2017, 2018\)](#), [Pariat et al. \(2017\)](#), [Park et al. \(2018\)](#), [Sharpe & Murray \(2017\)](#).

Appendix D

List of acronyms

2D	: Two-dimensional
3D	: Three-dimensional
ACC	: Accuracy
AD	: Appleman's discriminant
AIA	: Atmospheric Imaging Assembly
API	: Application programming interface
AR	: Active region
ASSA	: Automatic Solar Synoptic Analyzer
AUC	: Area under curve
BIAS	: Bias
BS	: Brier score
BSS	: Brier skill score
CCMC	: Community Coordinated Modeling Center
CME	: Coronal mass ejection

CNRS	: Centre national de la recherche scientifique
COSPAR	: Committee on Space Research
D-RAP	: D region absorption prediction
DB	: Data base
DEM	: Differential emission measure
DNA	: Deoxyribonucleic acid
DONKI	: Space weather database of notifications, knowledge, information
EASA	: European Aviation Safety Agency
EM	: Emission measure
ESA	: European Space Agency
ESWACC	: European Space Weather Assessment and Consolidation Committee
ETS	: Equitable threat score
EU	: European Union
EUV	: Extreme ultraviolet
FAR	: False alarm ratio
FHNW	: Fachhochschule Nordwestschweiz
FITS	: Flexible image transport system
FLARECAST	: Flare likelihood and region eruption forecasting
FN	: False negative
FP	: False positive
FOV	: Field of view
GAIA	: Gaussian AIA
GNSS	: Global Navigation Satellite System
GOES	: Geostationary Operational Environmental Satellites
GPS	: Global Positioning System
HARP	: HMI active region patch
HEK	: Heliophysics Events Knowledgebase
HELCATS	: Heliospheric cataloguing, Analysis and Techniques Service
HMI	: Helioseismic and Magnetic Imager
HSS	: Heidke skill score
HTTP	: Hypertext Transfer Protocol
IDL	: Interactive data language
JSOC	: Joint Science Operations Center
JSON	: Javascript object notation
LASCO	: Large Angle and Spectrometric Coronagraph
LASSO	: Least absolute shrinkage and selection operator
LOS	: Line of sight
LOWCAT	: Low corona catalog
MEDOC	: Multi Experiment Data & Operation Center
MHD	: Magnetohydrodynamics
MLP	: Multi-Layer Perceptron
MOSWOC	: Met Office Space Weather Operations Centre
NaN	: Not a number
NASA	: National Aeronautics and Space Administration
NetDRMS	: Network data record and management system
NOAA	: National Oceanic and Atmospheric Administration
NPDA	: Nonparametric discriminant analysis
NRT	: Near-realtime
O2R	: Operations to Research
OR	: Odds ratio

ORSS	: Odds ratio skill score	SOHO	: Solar and Heliospheric Observatory
PFE	: Potential field extrapolation	SRS	: Solar region summary
PIL	: Polarity inversion line	SSA	: Space Situational Awareness
POD	: Probability of detection	STEREO	: Solar Terrestrial Relations Observatory
POFD	: Probability of false detection	SVM	: Support vector machine
PostgreSQL	: Postgres structured query language	SWPC	: Space Weather Prediction Center
PROTEC	: Protection of our assets in space	SXI	: Solar X-Ray Imager
R2O	: Research to Operations	TEC	: Total electron content
RF	: Random forests	TN	: True negative
RNN	: Recurrent neural network	TP	: True positive
ROC	: Receiver operating characteristic	TS	: Threat score
SDO	: Solar Dynamics Observatory	TSS	: True skill statistic
SEDI	: Symmetric extremal dependence index	URL	: Uniform resource locator
SEP	: Solar energetic particles	WCS	: World Coordinate System
SHARP	: Space Weather HMI active region patch	WP	: Work package
SOC	: Self-organized criticality		

Cite this article as: Georgoulis MK, Bloomfield DS, Piana M, Massone AM, Soldati M, et al. 2021. The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era. *J. Space Weather Space Clim.* **11**, 39. <https://doi.org/10.1051/swsc/2021023>.