Gabriele Pierantoni
Eoin Carley
Jason Byrne
David Perez-Suarez
Peter T. Gallagher

A WORKFLOW-ORIENTED APPROACH TO PROPAGATION MODELS IN HELIOPHYSICS

Abstract

The Sun is responsible for the eruption of billions of tons of plasma and the generation of near light-speed particles that propagate throughout the solar system and beyond. If directed towards Earth, these events can be damaging to our tecnological infrastructure. Hence there is an effort to understand the cause of the eruptive events and how they propagate from Sun to Earth. However, the physics governing their propagation is not well understood, so there is a need to develop a theoretical description of their propagation, known as a Propagation Model, in order to predict when they may impact Earth. It is often difficult to define a single propagation model that correctly describes the physics of solar eruptive events, and even more difficult to implement models capable of catering for all these complexities and to validate them using real observational data.

In this paper, we envisage that workflows offer both a theoretical and practical framework for a novel approach to propagation models. We define a mathematical framework that aims at encompassing the different modalities with which workflows can be used, and provide a set of generic building blocks written in the TAVERNA workflow language that users can use to build their own propagation models. Finally we test both the theoretical model and the composite building blocks of the workflow with a real Science Use Case that was discussed during the 4th CDAW (Coordinated Data Analysis Workshop) event held by the HELIO project. We show that generic workflow building blocks can be used to construct a propagation model that succesfully describes the transit of solar eruptive events toward Earth and predict a correct Earth-impact time

1. Introduction

On a daily basis, the Sun produces spectacular eruptions of plasma, magnetic field and high energy particles, known as coronal mass ejections (CMEs; see Fig. 1). CMEs carry large amounts of energy through the solar system to interact with the planets in a manner referred to as 'space weather' [28].
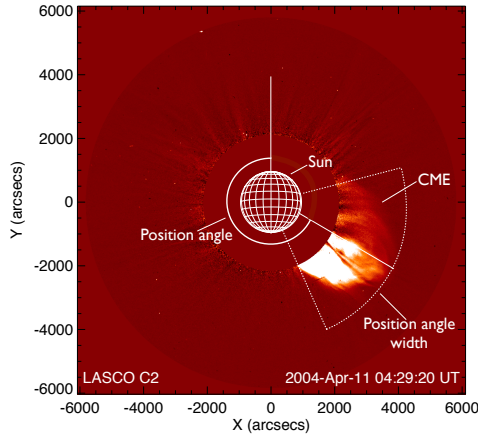

Figure 1. A coronal mass ejection observed by the LASCO C2 coronagraph on 2004-April-11. The Sun is represented by the latitude and longitude grid at image center. Also shown is the CME position angle, which is the position of the CME in degrees clockwise from solar north. The position angle width is also indicated. CMEs are the most energetic eruptions in the solar system and can travel toward Earth in excess of $2000 \, \text{km s}^{-1}$.

They are the main driver of the spectacular aurorae often observed at Earth's poles, but can be of detriment to our technological infrastructure as they interfere with satellite operations, telecommunication and GPS networks, and cause radiation increases for polar air travel and manned spaceflight. To this end, many scientifically-motivated observations of the Sun and CMEs are also useful in a real time application, to detect and alert us of a CME's occurrence. An advance-warning system can allow us to undertake preventative measures to minimise the damage of a potentially geo-effective impact at Earth; but this type of system requires an inherent understanding of the physics governing CME propagation in the solar wind[1], and their evolution through the interplanetary environment, so that an estimated time of arrival (ETA) may be produced.

In Heliophysics, Propagation Models are used to describe the transit of CMEs and high energy particles from the Sun towards the planets and predict an ETA of these potentially hazardous events. These models provide a computational framework

---

[1]A constant stream of plasma from the Sun towards the planets

to describe the great complexity and variety of different solar eruptive events and their interaction with the environment of the solar system. Although they are fundamental to most investigations in heliophysics, the models can be difficult to define, implement and use.

One major problem with Propagation Models is the need to define the parameters that govern the model, for example the model input may require speed and direction of the eruption. Such information is not always available, forcing the model user to perform wild guesses and validate her/his parameter assumptions after the model is executed.

A first solution to this problem was proposed in [1] whereby a simple ballistic model (SHEBA) was extended so that the user was required only to guess a range for every parameter allowing the model to be executed as a parameter sweep job. Once the results where obtained for each value of the selected parameter range, a ranking function is used to decide which parameter value is the correct one e.g., which speed correctly describes the propagation of the event. We termed this approach Advanced Propagation Model (APM) where paramer sweep jobs are executed and the results validated by comparing the outcomes with observational data.

Although this approach has the merit of freeing the users from the need of guessing parameters and manually validating the results, it still has many drawbacks.

Firstly, it is not flexible enough to accomodate the large variety of usage patterns that are required by the community: more specifically, there are many different ways in which events can be discovered, parameters values and ranges inferred from events, and results validated. Secondly, the implementation of the prototype APM proved to be too rigid to accomodate many use cases that were of real scientific relevance. In order to overcome these shortcomings, we propose an improved mathematical framework that caters for different scenarios, and a workflow-based [10, 20, 15, 12, 25] approach that ensures the flexibility required to describe the variety of solar eruptive events and the range of properties (speeds, locations) that these events may have.

This paper is structured as follows:

- *Heliophysics Propagation Models and Observations*, in Section 2, introduces the use of Propagation Models in Heliophysics and relates them to the available observation data;
- *Mathematical Formalism for Propagation Models*, in Section 3, describes the Mathematical Formalism we use to describe the propagation models;
- *Scientific Use Case*, in Section 4, details the scientific use case that we use to validate the model and the implementation;
- *Workflows for Propagation Models in Heliophysics*, in Section 5, introduces the building blocks that have been developed to implement the Science Use Case;
- *Conclusions and Future Work*, in Section 6, concludes the paper and introduces future directions of research.

2. Heliophysics Propagation Models and observations

This paper will primarily use a combination of Propagation Models and Heliophysics observational data. Observational data of CMEs have been compiled into catalogs that aid CME forecasting and help provide realtime alerts to the space weather community. Similarly a variety of theoretical CME and solar wind models have been built to take input from the observational data and/or in-situ measurements, and simulate CME propagation through the relevant portions of the solar system. Predicting the ETA of a CME therefore requires the use of observational data, an understanding of CME and solar wind propagation theory, and the computational resources necessary to promptly model and predict their evolution through space. Observational data exists in the form of telescopic imagery from such missions as the Solar & Heliospheric Observatory [11] launched in 1995, and the more recent Solar Dynamics Observatory [16] launched in 2010. From the imaging data provided by such missions, various CME properties may be determined, such as the start-time and location, and early-stage kinematics and morphology. These properties are then used in a propagation theory that normally considers some form of ballistic motion of the CME on top of a simplified magnetohydrodynamic (MHD) treatment of its evolution and interaction with the solar wind. Details on the current CME models in use/development may be found in, e.g., ENLIL [13], HAFv2, ISPM, or STOA [18], with each determining its own predicted arrival time of the bulk CME material and/or associated shock wave.

Perhaps the most dominant physical aspect in propagation models is the effects of solar wind drag that can act on CMEs; with the general result that fast CMEs are slowed down, and slow CMEs sped up, to the surrounding solar wind speed [17]. Different models generally quantify this effect in different ways, and can therefore produce varying predictions on CME propagation and evolution. *In-situ data*[2] is often used as a validity check on the results of a CME model run, by comparing the predicted and observed arrival time of a CME and/or shock, and examining the parameters that may be responsible for any sources of error. Thus the model parameters may be iterated to an input that most appropriately simulates the true CME evolution, and the model itself adjusted for improved future performance.

Performing such iterations over multiple case-studies may be made substantially easier through the introduction of a workflow type process. This is especially true when one considers that the SOHO/LASCO instrument has observed $> 10^4$ CMEs since launch, and extensive catalogs built from the resulting metadata (compiled by both human users and automated algorithms) have been produced. Furthermore, there exists a wealth of similar catalogs of other solar phenomena, such as flares, filaments, coronal holes, and shocks. In order to utilise all such event information in tandem, we use the Helio Event Catalog (HEC) that is a compilation of all solar

---

[2]Data pertaining to the local environment at the space location

activity catalogs. The HEC[3], as well as other services such as the Propagation Model are a set of workflows legacy of the HELIO Project [5, 21].

Our approach is to construct workflows that use the vast amounts of metadata in the HEC to construct propagation models of CMEs. This allows us to test the physical validities of CME propagation theory and check the reliability of the metadata for use in CME physics.

3. Mathematical formalism for Propagation Models

A *Propagation Model* can be seen as a mathematical expression that describes the propagation of a physical phenomena. In Heliophysics, propagation models are used to describe the propagation of different phenomena throughout the Solar System. As Propagation Models rely on parameters that cannot be known exactly, they usually return an estimated range for the *Expected Arrival Time* rather than a single value.

3.1. Simple Propagation Model

A simple Propagation Model is described in equations (1) and (2) and in Figure 2

$$eta_t = P(\overrightarrow{x_s}, \overrightarrow{p}, t_s) \tag{1}$$

$$eta = \{t_{\min}, t_{\max}\} \tag{2}$$

where eta_t is the expected arrival time at target, $P(\dots)$ is the function that models the propagation of the phenomena, $\overrightarrow{x_s}$ is the location of the source of the event, $\overrightarrow{p}$ is the set of parameters that govern the model, and t_s is the time at which the phenomena left the source.
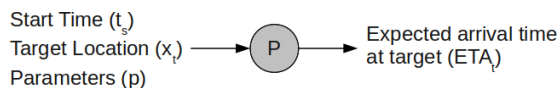

Figure 2. Conceptual Representation of a Simple Propagation Model.

This basic approach to a propagation model has two main weaknesses. Firstly, it is difficult to infer the parameters $\overrightarrow{p}$ that govern the propagation and, secondly, it may be difficult to validate the results. To overcome such difficulties, we proposed the concept of an *Advanced Propagation Model* in [22] that allowed for automatic estimation of parameters and validation of results. In this paper we expand and modify that approach to allow a more flexible implementation. We envisage the following scenarios:

- *Simple Propagation Model* to model the propagation of a single event with parameter estimation and result validation performed manually by the user
- *Assisted Propagation Model* to assist the user in defining the parameters

---

[3]`http://hec.helio-vo.eu/hec/hec_gui.php`

- *Validated Propagation Model* to assist the user in the validation of results
- *Iterative Propagation Model* to allow for the iteration of the propagation model to incrementally increase the precision of the results

3.2. Assisted Propagation Models

An *Assisted Propagation Model* (see Fig. 3), is an extension of the *Simple Propagation Model* where a special function allows the user to evaluate some or all the parameters.
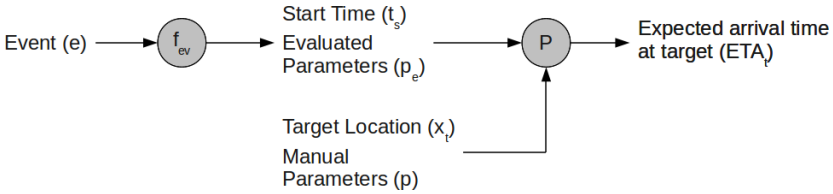
Figure 3. Conceptual Representation of an Assisted Propagation Model.

$$eta_t = P(\vec{x_s}, \vec{p_m}, \vec{p_e}, t_s) \tag{3}$$
$$\vec{p_e} = f_{ev}(e) \tag{4}$$

where f_{ev} is the function that evaluates a subset of the parameters $\vec{p_e}$ from event e, the subset of paremeters that cannot be evaluated by f_{ev}, $\vec{p_m}$ is still set manually.

$$\vec{p_m} \subseteq \vec{p} \tag{5}$$
$$\vec{p_e} \subseteq \vec{p} \tag{6}$$
$$\vec{p_e} \bigcup \vec{p_m} \equiv \vec{p} \tag{7}$$
$$\vec{p_e} \bigcap \vec{p_m} \equiv \emptyset \tag{8}$$

Equation 3 can be extended to allow automated searches for events in catalogs as describe in Figure 4 by defining a *query function* that uses catalogs to automatically find events given a time range.
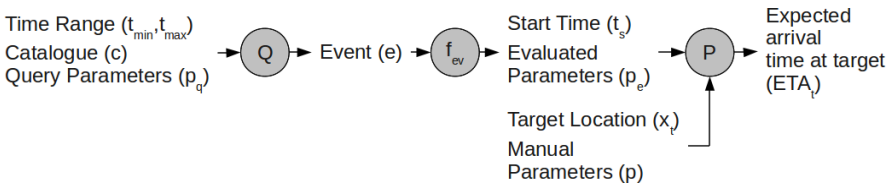
Figure 4. Conceptual Representation of an Assisted Propagation Model with Search of Events.

$$e = Q(t_{\min}, t_{\max}, c, p_q) \tag{9}$$

where $t_{\min}$, $t_{\max}$ is the timerange of the query, c is the *catalog* to be used and $\overrightarrow{p_q}$ are the set of extra parameters that govern the execution of the query. By combining (9), (4) and (3) we obtain the full equation for an assisted propagation model.

$$eta_t = P(\overrightarrow{x_s}, \overrightarrow{p_m}, Q(t_{\min}, t_{\max}, c, \overrightarrow{p_q}), t_s) \tag{10}$$

This approach can be extended to analyze multiple events within a time range to support statistical analysis of similar phenomena (see Fig. 5).
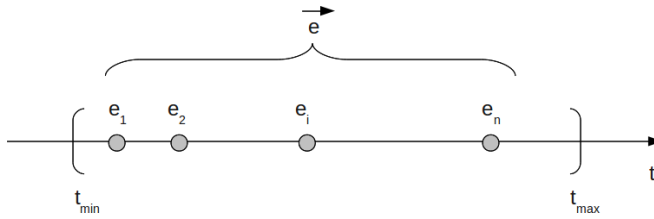

Figure 5. Query for all events in a time range.

$$\overrightarrow{e} = Q(t_{\min}, t_{\max}, c, p_q) \tag{11}$$

where $\overrightarrow{e}$ is the sequence of events found within time range $t_{\min}, t_{\max}$ on catalog c.

By combininig 11 with 10, we obtain the model for assistend analysis of multiple phenomena.

$$\overrightarrow{eta_t} = P(\overrightarrow{x_s}, \overrightarrow{p_m}, Q(t_{\min}, t_{\max}, c, \overrightarrow{p_q}), t_{s_i}) \tag{12}$$

3.3. Validated Propagation Models

As parameters governing the model can be infered either manually or by functions, it is important to assist the user in validating the result as described in Figure 6.
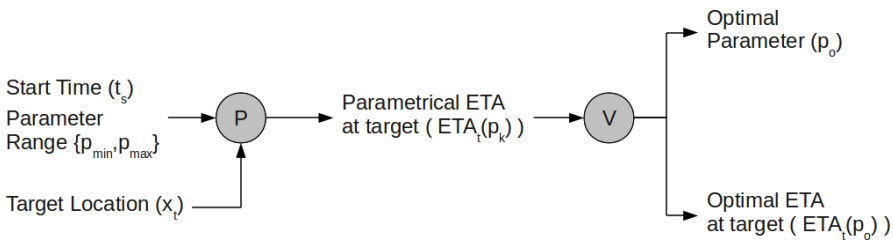

Figure 6. Conceptual Representation of a Validated Propagation Model.

This can be combined by executing the propagation models for a range of parameters as a paremeter sweep job, validate the results and infer the correct parameter *a posteriori*, instead than *a priori*.

A first approach to this problem has been proposed in [22], we further expand here the concept of validation to foster flexibility. First we define validation functions that can be either boolean or return a ranking value. The first are used to simply accept or reject a result, the second to find optimal set of parameters.

$$eta_t(p_k) = P(\vec{x_s}, \overrightarrow{p_m}, \overrightarrow{p_k}) \tag{13}$$

$$p_{k_{\min}}, p_{k_{\max}} = f_{ev}(e) \qquad \forall k \in k_{\min}, k_{\max} \tag{14}$$

where p_k is sequence of parameters within a range $k_{\min}, k_{\max}$ infererred from the events using a modified version of the evaluator function $f_{ev}(\dots)$ (4). Then we define validation and ranking functions used to assess which was the best parameter.

$$v = V(eta_t(p_k)) \qquad \forall v \in true, false \tag{15}$$

$$r = R(eta_t(p_k)) \qquad \forall r \in N \tag{16}$$

The validation V and ranking R are used to find the optimal parameter p_o as in

$$p_o \: : f_v(p_o) = true \tag{17}$$

$$p_o : \max_{p_{\min} \leq p_i \leq p_{\max}} (f_r(p_i)) \tag{18}$$

3.4. Iterative Propagation Models

Finally, it is necessary to model *iterative* models whereby a model or different models are iteratively invoked until a certain condition is obtained. So, we define an iterative propagation model that we will later implement in workflows.

$$eta_{t_n} = f_{pm}(\vec{x_s}, \overrightarrow{p_{m_{n-1}}}, \overrightarrow{p_{k_{n-1}}}) \tag{19}$$

$$\overrightarrow{p_{k_n}} = f_{ev}(eta_{t_n}) \tag{20}$$

where eta_{t_n} is the expected arrival time at the nth iteration of the model and $\overrightarrow{p_{k_n}}$ is the nth refinement of a subset of parementers that can be inferred by (20) and $\overrightarrow{p_{m_n}}$ is the nth refinement of the subset of parementers that must be set manually.

4. Scientific use case

CMEs may travel with a speed in excess of $2000 \, \text{km s}^{-1}$ [29] and often drive shock waves through the solar atmosphere and solar wind [27, 19]. Given that these shocks can be responsible for radio communication blackouts and the production of potentially hazardous high energy particles, there is a concetrated effort in space weather studies to understand their nature and their relationship to CMEs.

The shocks have a variety of physical manifestations, including high intensity bursts of radio radiation that are regularly observed from ground and space-based observatories. Separately, shocks may also be detected in-situ by spacecraft located at Earth and beyond [24]. The spacecraft monitor the local density, velocity and

temperature of their environment, and the arrival of a shock is usually identified by a sudden increase in all of these properties. However, despite the fact that shocks may be readily identified using both radio and in-situ data, little work has been done on the relationship between these observables i.e., if a shock is detected at radio wavelengths, will the same shock also be detected in-situ by a spacecraft at Earth or beyond [9]?

There are a number of catalogs that exist in the HEC that allow us to investigate the relationship between radio detection and the in-situ detection of CME-driven shocks. One such catalog utilised in this study includes shock radio burst detections from the WAVES instrument [7] on board the WIND spacecraft located at Earth. Each of the radio bursts in the catalog may be associated with a CME observed by the LASCO coronagraphs and listed in the appropriate catalog in the HEC. The CME properties from this catalog may then be used to run a propagation model (thus extending it to be an *Assisted Propagation Model*) and estimate an ETA at a desired location/spacecraft (since the shock is physically located in front of the CME, the shock should arrive just before or at the expected ETA of the CME). A catalog of in-situ shocks may then be searched for a positive shock detection at the ETA calculated from the propagation model. Hence, the goal of our science case is to show that a shock observed at radio wavelengths had a counterpart observed in-situ, with the two being linked by the propagation of a CME through the solar system. From a mathematical point of view, with this science case, we validated the *Assisted* and *Iterative* models.

5. Workflows for Propagation Models in Heliophysics

There are a number of existing workflows that have been built by the heliophysics community [6] that aid the analysis of solar physics data – these can be found at `http://www.myexperiment.org/groups/101.html`. The workflows involve extracting metadata from catalogs of CMEs, solar flares, solar wind properties or shocks. They range in complexity from some very basic event catalog queries to workflows that produce results of scientific merit. For this study, we systematically assessed the function of these workflows and categorised them in terms of functionality, usefulness for our purposes and type of workflow. Amongst the types of workflow we identified were simple catalog queries (6 workflows), workflows that produce results of scientific merit (12), and simple utilities and tools (21) e.g., for parsing data or sorting arrays and lists.

The goal of this study was to use existing workflows and new workflows that may be used as generic building blocks to produce a master workflow. The generic blocks should be flexibile and usable alongisde other components to produce a desired result. To this end, there were a number of workflows we identified for such a purpose, some of which were built for the 4th HELIO workshop held at Trinity College Dublin, Ireland during September 2012. The workflow components that we chose were ones that implement the SHEBA propagation model (myExperiment ID 2764), retrieve

a variety of solar wind properties using the ACE spacecraft (myExperiment ID 3265), and a small utility workflow for parsing XML sheets (myExperiment ID 1911).

We added to the exhisting workflows, simple and re-usable components to query solar event catalogs using web services; these were constructed in such a way to allow the user tailor the query e.g., query for a certain number of results or to order the results by a particular paramater. The generic workflow components used to construct our workflow are given in Table 1.

Table 1

Workflow components used in the implementation.

myExperiment ID	Purpose	Workflow Type
2764	Run SHEBA	Scientific
3265	Average solar wind speed from ACE	Scientific
1991	Parse XML Sheet	Utility/Tool
3982	Catalog Query (parameters)	Query web service
3983	Catalog Query (parameters, order, limit)	Query web service
3983	Catalog Query (with SQL condition)	Query web service

5.1. Implementation of the workflow

For the specific science case of analysing CME-driven shocks the basic building blocks of the workflow should consist of the following

- Querying particular event catalogs and receiving metadata in return. There are a number of workflow inputs that allow the user to choose the parameters to be returned, limit and order the search results etc. This component of the workflow is generic and should give the ability to query any catalog for any desired parameter.
- Parsing the metadata for the desired information (usually lists of parameters such as event times and velocities, for example). Again, this is a generic and re-usable component that should parse any metadata in XML format and return a list of the desired parameters.
- Implementation of the the SHEBA propagation model. The input parameters for this are from the lists retrieved from the query and parsing of the metadata.

The initial query is to the *wind_typeii_soho_cme* catalog, this contains information on all CME-associated shock radio burst events detected by the WIND/WAVES instrument. The parameters queried are the start time and end time of the event, the location on the solar surface, CME velocity, and the strength of the X-rays from the associated solar flare; the workflow component that allows such a catalog query is shown in Figure 9 and its placement in a conceptual workflow is shown as component Q1 in Figure 7. This component returns an XML sheet of all desired parameters which is then parsed to produce lists of velocity and times etc. The events are ordered by velocity so that, of all events returned, the fastest CME may be chosen. The time of the event is then passed to a query of the *soho_lasco_cme* catalog (component Q2
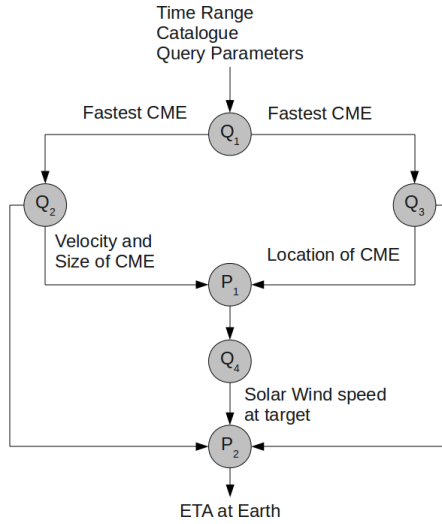
Figure 7. Conceptual workflow showing the placement of queries (Q) and the implementation of the propagation model (P).

in Figure 9), from which a number of CME parameters are chosen, including position angle, angular width, final velocity and velocity uncertainty. The time of the event is also passed to a component that queries the *goes_sxr_flare* catalog and obtains the latitude, longitude and strength of the associated solar X-ray flare (component Q3 in Figure 9).

The CME speed, uncertainty on the speed, width, time and starting position are then passed to the SHEBA propagation model from which an ETA at Earth is produced, indicated in Figure 11 and shown as P1 in Figure 9. Since the CME is unlikely to propagate ballisticly, this initial ETA is unlikely to be the actual arrival time. We need to account for the fact that the CME will be slowed by the solar wind due to aerodynamic drag. This is done by querying a catalog of solar wind speeds at the ACE spacecraft (located at Earth) at the initial ETA we obtained from the propagation model, this workflow component is shown in Figure 10 and as component Q4 in Figure 9. The solar wind speed is then used to re-define a new and decreased CME speed to be used in a second run of the propagation model (see Figure 11 and P2 in Figure 9); the ETA of this second iteration is then compared to a catalog of in-situ shocks and CME detections in order to confirm if there was any positive shock detections.

5.2. Results

As a test case for our workflow and propagation we chose to analyse an eruptive event that occurred on 11 April 2004. The event was associated with a X-ray flare and a CME which was observed by the LASCO telescopes at 04:30 UT with central
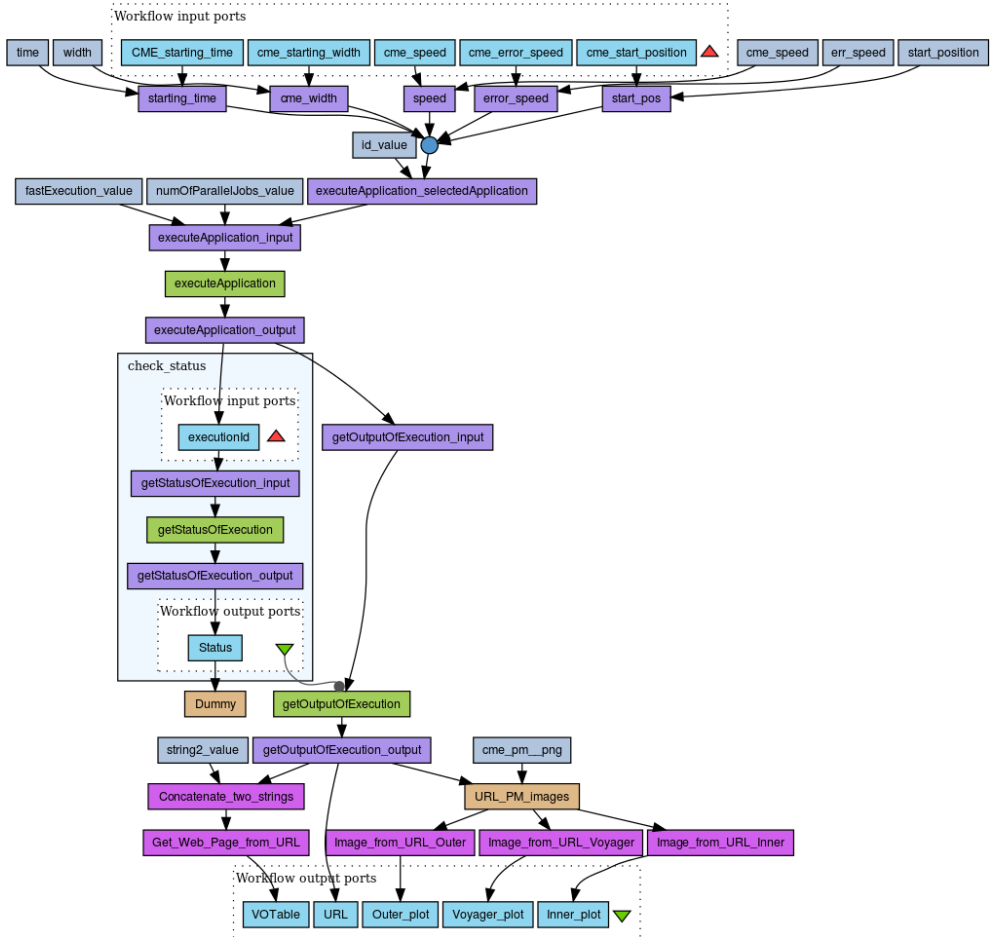
Figure 8. Implementation of the SHEBA propagation Model.

position angle 203°, angular width 314°, and speed 1645 km s^{-1}. The event was associated with a radio burst from a solar shock wave observed by the WIND/WAVES instrument at 04:20 UT. The *test case* inputs/outputs were specified as follows:

1. A time interval is specified and input to the WIND/WAVES shock radio bursts and CMEs list to retrieve a list of events within the given time-range of interest.
   Time range: 2004/04/01 00:00:00 – 2004/04/30 00:00:00 UT
2. The list of candidate events during this month long period were ranked in order of decreasing CME speed, with the intent that the single fastest event in the *test case* sample was chosen, as described above.
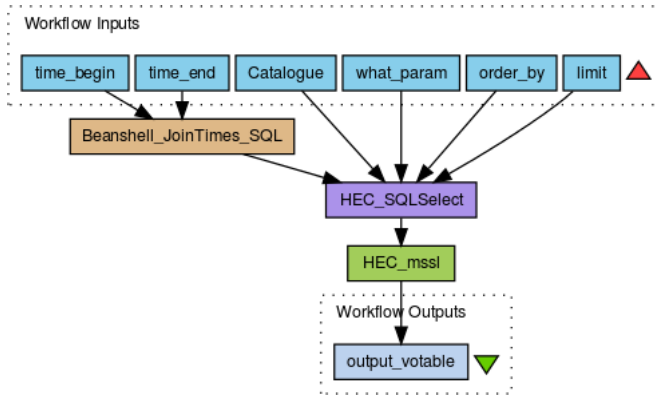
Figure 9. Generic workflow to query the Helio Event Catalog (HEC) for any catalog and any parameters of that catalog. The inputs parameters include the desired time range, parameters to be returned, parameter to order by and a limit on the search result.

Shock radio burst	*Time range: 2004/04/11 04:20 – 05:35 UT*
	Observed frequency: 14000 – 500 kHz
Flare	*Location on the solar surface: South-14 West-47*
	X-ray Class: C9.6
CME	*Start time: 2004/04/11 04:30 UT*
	Central position angle: 203°
	Angular width: 314°
	Speed: 1645 km s^{-1}

3. The GOES [4] Soft X-ray Flare List[5] was then inspected for any associated flaring activity of the relevant class, within a specified window of ± 1 hour on the start time of the shock radio burst, to obtain the catalogued source longitude on disk.
   Time range: 2004/04/11 03:20 – 05:20 UT
   Time of maximum X-ray intensity (t_{start}): 04:19 UT
   Longitude on solar surface (λ_{lon}): 46°

4. The LASCO CME Catalog[6] is inspected in order to associate CME parameters from the relevant detection in the time range of the shock radio burst. In this case the necessary parameters are the CME initial and final speeds, and angular width. The choice of catalog can be changed, for example to call one of the automated CME catalogs such as CACTus.
   v_{init} :   *1953 km s^{-1}*
   v_{final} :   *1340 km s^{-1}*
   θ_{cme} :   *314°*

---

[4]Geostationary Operational Enviromental Satellite
[5]http://www.ngdc.noaa.gov/stp/solar/solarflares.html
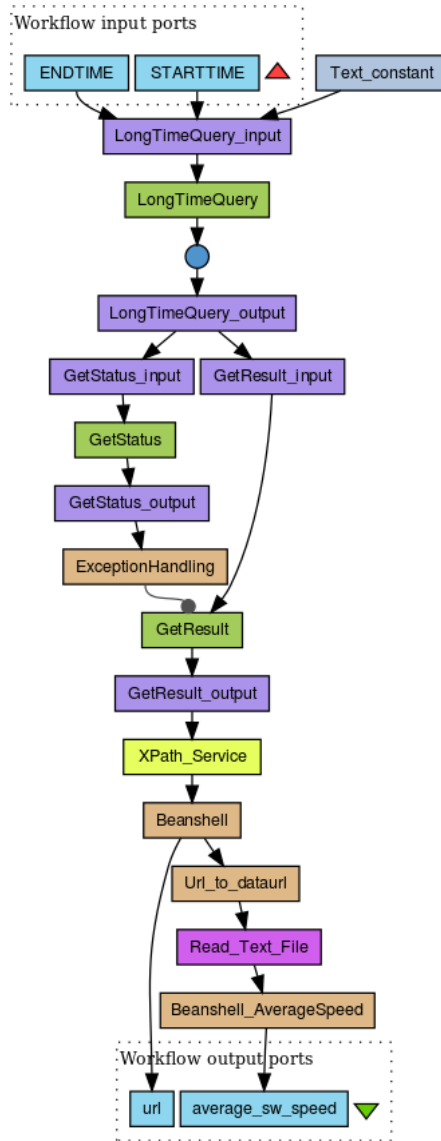[6]http://cdaw.gsfc.nasa.gov/CME_list/

Figure 10. Workflow component that queries a catloague of solar wind speeds detected by the ACE spacecraft, located at Earth. The workflow takes an input of a time range outputs the average solar wind speed within this time range. Typical solar wind speeds tend to be anywhere in the range of 400–$800\,\mathrm{km\,s^{-1}}$. If a CME differs in speed from thw wind speed, it experiences an aerodynamic drag.
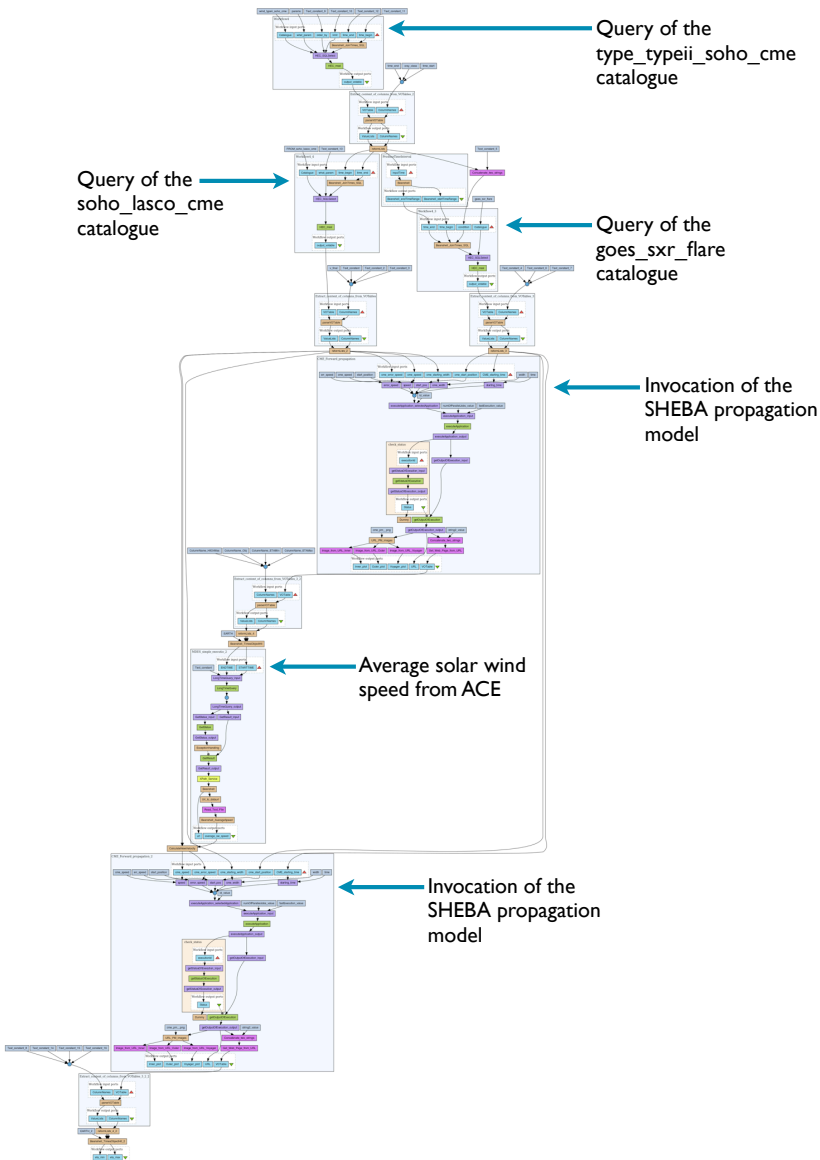
Figure 11. Overal workflow including queries to various catalogs and implementation of the SHEBA propagation model. The initial query is to the *wind_waves_typeii_cme* catalog, where the fastest CME in a time range is chosen. The paramaters of this CME and the associated solar flare are then obtained fom the *soho_lasco_cme* and *goes_sxr_flare* catalogs, respectively. The metadata from these three catalogs is then used to run the SHEBA propagation model. The model is run a second time, with a re-calibrated CME speed by taking into account the effects of drag due to the solar wind.

5. The CME speed is determined as $v_{cme} = v_{final} \pm \sigma_v$ where an initial estimate of the uncertainty on the CME speed is calculated as $\sigma_v = \frac{|v_{final} - v_{init}|}{2}$.

   v_{cme} :   *1340 ± 306 km s^{-1}*

6. SHEBA is run with the following input parameters: CME start time from the peak time of the associated flare t_{start}; trajectory from the associated the longitude on the solar surface λ_{lon}; speed v_{cme}, and angular width θ_{cme}.

7. From the ballistic CME model, an expected timeframe of arrival at Earth is determined. If an event is not deemed Earth-directed it is flagged as so. The in-situ data from the ACE spacecraft is queried via the Automated Multi Dataset Analysis web service[7], and an average speed of the solar wind $\bar{v}_{sw}$ during this timeframe is calculated.

   ETA range: 2004/04/12 05:57:54 – 21:10:41 UT
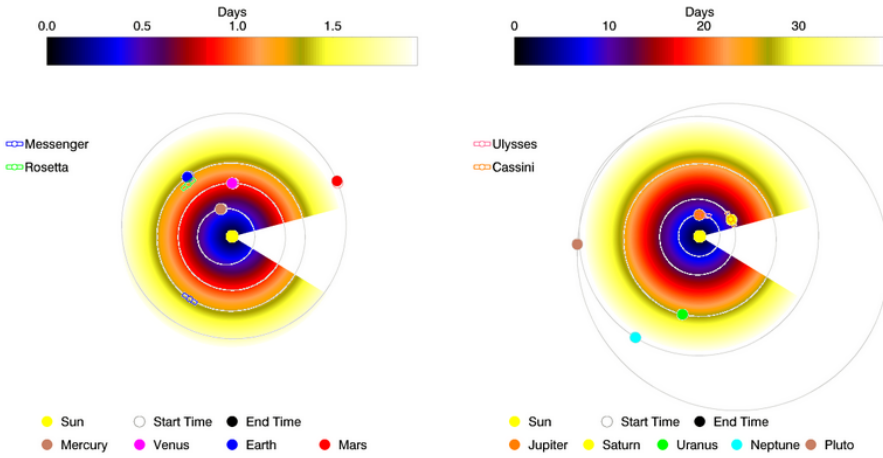 $\bar{v}_{sw}$:   *442 km s^{-1}*

Figure 12. The output of the SHEBA propagation model, showing the expected arrival time of the CME at the solar system planets in days after 2004-04-11 04:19 UT. In this instance the CME is unphysically wide, owing to an erroneous paramater in the *soho_lasco_cme*. However, this does not invalidate the results, given that there was an arrival of a shock at Earth at the expected time.

8. From the average solar wind speed, a new velocity of the CME is calculated to essentially account somewhat for the influence of drag. The average solar wind speed is used to modify the input CME speed by lowering the uncertainty interval to match it as the lower bound (or raise it to the upper bound as the case may be, though unlikely for the fastest CME). While the upper bound is kept fixed, the modified CME speed between the bounds is calculated as $v'_{cme} = \frac{1}{2}\left(\bar{v}_{sw} + \frac{v_{final} + v_{init}}{2}\right)$ with new uncertainty

---

[7]http://manunja.cesr.fr/Amda-Helio/

$\sigma'_v = \frac{1}{2}(\sigma_v + v_{final} - \bar{v}_{sw})$. These are used to rerun the ballistic CME propagation model. $v'_{cme}:$ *1044 ± 602 km s^{-1}*

9. The predicted impact timeframes of the CME at the relevant locations throughout the heliosphere may be output from the workflow, the results of the propagation model are shown in Figure 12. In this instance we chose an imapct time at Earth.

   Earth arrival timeframe: *2004/04/12 05:57 – 2004/04/15 03:47 UT*

10. The final step is to check if an in-situ shock was detetced at Earth during calculated impact time. The catalog of in-situ shock detetctions to be parsed is from the CELIAS/MTOF/PM instrument [14] on board the SOHO spacecraft. During the ETA given by the workflow, a shock was detected in-situ at 2004/04/12 17:35 UT, see Figure 13. The positive identification of a shock at the expected time makes it likely that it is the same shock which produced the radio burst, observed at 2004/04/11 04:20 UT. With regard to our science case stated above, the workflow succesfully related a radio and in-situ shock detection via a CME propagation model.

6. Conclusion and future work

Adopting a workflow-centered approach has proved successful and advantageous on different counts.

Firstly, by developing basic building blocks that invoke services to query catalogs, extract metadata and execute propagation models we made publicly available to the community a set of re-usable and flexible components that can be used to build different sorts of Propagation Models thus overcoming the rigidity of the approach proposed in [22].

Secondly, we have tested the validity of this approach by building a complex workflow that models a propagation model that is both *Assisted* and *Iterative* as it extracts parameters from events and it runs twice, correcting the parameters of the second run with values of *in situ* data at a time obtained in the first run.

Finally, we have validated the implementation by investigating a real and significant CME event (April 11th 2004) and found consistency in the expected arrival time and shock wave signatures in *in situ* data at Earth.

Although this approach has already lead to interesting and relevant results, there is still great scope for improvement both from the technological and scientific point of view.

From the technological point of view, future work will focus on two main directions: to interface the system with the *HELIOGate* portal developed by the SCI-BUS project and to support workflow interoperability, an effort supported by the ER-FLOW project. From a scientific point of view, we will focus on extending the workflows to cather for automatic validation of the results thus obtaining a *Validated Propagation Model*.
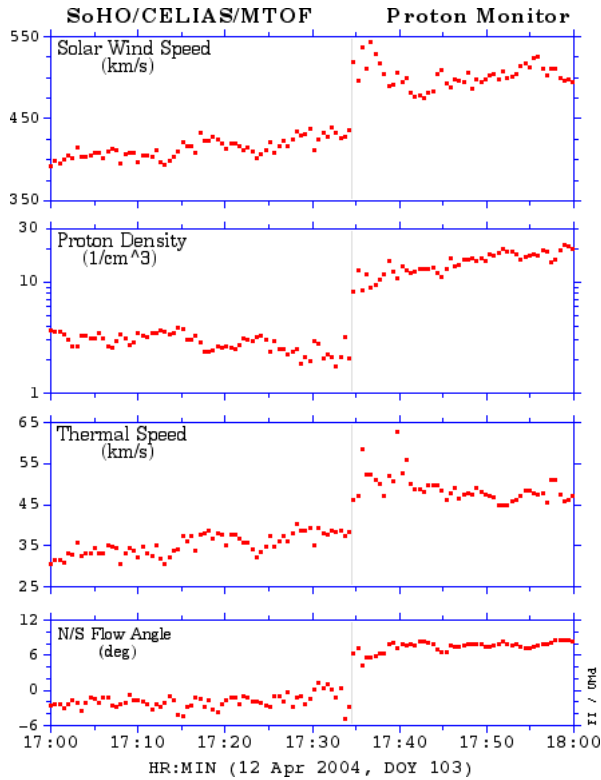
Figure 13. The in-situ detection of a shock with the CELIAS/MTOF Proton Montior onbaord the SOHO spacecraft. The grey line demarcates the arrival of a shock, where the speed, density, temperature and flow angle each experiences an abrupt change in value. *Image coutesy of the SOHO CELIAS team.*

Acknowledgements

References

[1] *ER-FLOW Project Page* – `http://www.http://www.erflow.eu/`.

[2] *Grid Ireland Web Page* – `https://grid.ie/`.

[3] *HELIO Project Page* – `http://www.helio-vo.eu/`.

[4] *SCI-BUS Project Page* – `https://www.sci-bus.eu/web/guest`.

[5] Bentley R., Brooke J., Csillaghy A., Fellows D., Blanc A. L., Messerotti M., Pérez-Suárez D., Pierantoni G., Soldati M.: HELIO: Discovery and analysis of data in heliophysics. *Future Generation Computer Systems*, **vol. 29**(8), pp. 2157–2168, 2013. ISSN 0167-739X. `http://dx.doi.org//10.1016/j.future.2013.04.006`.

[6] Blanc A., Brooke J., Fellows D., Soldati M., Prez-Surez D., Marassi A., Santin A.: *Workflows for Heliophysics. Journal of Grid Computing*, **vol. 11**(3), pp. 481–503, 2013. ISSN 1570-7873. `http://dx.doi.org/10.1007/s10723-013-9256-5`.

[7] Bougeret J. L., Kaiser M. L., Kellogg P. J., Manning R., Goetz K., Monson S. J., Monge N., Friel L., Meetre C. A., Perche C., Sitruk L., Hoang S.: Waves: The Radio and Plasma Wave Investigation on the Wind Spacecraft. *Space Science Reviews*, **vol. 71**, pp. 231–263, 1995. `http://dx.doi.org/10.1007/BF00751331`.

[8] Brueckner G. E., Howard R. A., Koomen M. J., Korendyke C. M., Michels D. J., Moses J. D., Socker D. G., Dere K. P., Lamy P. L., Llebaria A., Bout M. V., Schwenn R., Simnett G. M., Bedford D. K., Eyles C. J.: The Large Angle Spectroscopic Coronagraph (LASCO). *Solar Physics*, **vol. 162**, pp. 357–402, 1995. `http://dx.doi.org/10.1007/BF00733434`.

[9] Cremades H., St. Cyr O. C., Kaiser M. L.: A tool to improve space weather forecasts: Kilometric radio emissions from Wind/WAVES. *Space Weather*, **vol. 5**, S08001, 2007. `http://dx.doi.org/10.1029/2007SW000314`.

[10] Deelman E., Gannon D., Shields M., Taylor I.: Workflows and e-Science: An Overview of Workflow System Features and Capabilities. *Future Gener. Comput. Syst.*, **vol. 25**(5), pp. 528–540, 2009. ISSN 0167-739X.
`http://dx.doi.org/10.1016/j.future.2008.06.012`.

[11] Domingo V., Fleck B., Poland A. I.: The SOHO Mission: an Overview. *Solar Physics*, **vol. 162**, pp. 1–37, 1995. `http://dx.doi.org/10.1007/BF00733425`.

[12] Gil Y., Deelman E., Ellisman M., Fahringer T., Fox G., Gannon D., Goble C., Livny M., Moreau L., Myers J.: Examining the Challenges of Scientific Workflows. *Computer*, **vol. 40**(12), pp. 24–32, 2007. ISSN 0018-9162.
`http://dx.doi.org/10.1109/MC.2007.421`.

[13] Gressl C., Veronig A. M., Temmer M., Odstrčil D., Linker J.A., Mikić Z., Riley P.: Comparative Study of MHD Modeling of the Background Solar Wind. *Solar Physics*, 2013. `http://dx.doi.org/10.1007/s11207-013-0421-6`.

[14] Hovestadt D., Hilchenbach M., Bürgi A., Klecker B., Laeverenz P., Scholer M., Grünwaldt H., Axford W. I., Livi S., Marsch E., Wilken B., Winterhoff H. P., Ipavich F. M., Bedini P., Coplan M. A., Galvin A.B., Gloeckler G., Bochsler P., Balsiger H., Fischer J., Geiss J., Kallenbach R., Wurz P., Reiche K. U., Gliem F.,

Judge D. L., Ogawa H. S., Hsieh K. C., Möbius E., Lee M. A., Managadze G. G., Verigin M. I., Neugebauer M.: CELIAS – Charge, Element and Isotope Analysis System for SOHO. *Solar Physics*, **vol. 162**, pp. 441–481, 1995.
http://dx.doi.org/10.1007/BF00733436.

[15] Hull D., Wolstencroft K., Stevens R., Goble C., Pocock M. R., Li P., Oinn T.: Taverna: a tool for building and running workflows of services. *Nucleic Acids Research*, **vol. 34**(suppl 2), pp. W729–W732, 2006.
http://dx.doi.org/10.1093/nar/gkl320.

[16] Lemen J. R., Title A. M., Akin D. J., Boerner P. F., Chou C., Drake J. F., Duncan D. W., Edwards C. G., Friedlaender F. M., Heyman G. F., Hurlburt N. E., Katz N. L., Kushner G. D., Levay M., Lindgren R. W., Mathur D. P., McFeaters E. L., Mitchell S., Rehse R. A., Schrijver C. J., Springer L. A., Stern R. A., Tarbell T. D., Wuelser J. P., Wolfson C. J., Yanari C., Bookbinder J. A., Cheimets P. N., Caldwell D., Deluca E. E., Gates R., Golub L., Park S., Podgorski W. A., Bush R. I., Scherrer P. H., Gummin M. A., Smith P., Auker G., Jerram P., Pool P., Soufli R., Windt D. L., Beardsley S., Clapp M., Lang J., Waltham N.: The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Physics*, **vol. 275**, pp. 17–40, 2012. http://dx.doi.org/10.1007/s11207-011-9776-8.

[17] Maloney S. A., Gallagher P. T., McAteer R. T. J.: Reconstructing the 3-D Trajectories of CMEs in the Inner Heliosphere. *Solar Physics*, **vol. 256**, pp. 149–166, 2009. http://dx.doi.org/10.1007/s11207-009-9364-3.

[18] McKenna-Lawlor S. M. P., Dryer M., Kartalev M. D., Smith Z., Fry C. D., Sun W., Deehr C. S., Kecskemety K., Kudela K.: Near real-time predictions of the arrival at Earth of flare-related shocks during Solar Cycle 23. *Journal of Geophysical Research (Space Physics)*, **vol. 111**, A11103, 2006.
http://dx.doi.org/10.1029/2005JA011162.

[19] Nelson G. J., Melrose D.B.: *Type II bursts*, pp. 333–359. 1985.

[20] Oinn T., Addis M., Ferris J., Marvin D., Senger M., Greenwood M., Carver T., Glover K., Pocock M. R., Wipat A., Li P.: Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics*, **vol. 20**(17), pp. 3045–3054, 2004. http://dx.doi.org/10.1093/bioinformatics/bth361.

[21] Pierantoni G., Coghlan B., Kenny E.: The Architecture of HELIO. In: *Krakow Grid Workshop*. 2010.

[22] Pierantoni G., Coghlan B., Kenny E., Gallagher P., Perez-Suarez D.: Extending the SHEBA Propagation Model to reduce Parameter-Related Uncertainties. *Computer Science*, **vol. 14(2)**, pp. 253–272, 2012.

[23] Pulupa M., Bale S. D.: Structure on Interplanetary Shock Fronts: Type II Radio Burst Source Regions. *Astrohpysicsl Journal*, **vol. 676**, pp. 1330–1337, 2008.
http://dx.doi.org/10.1086/526405.

[24] Pulupa M. P., Bale S. D., Kasper J. C.: Langmuir waves upstream of interplanetary shocks: Dependence on shock and plasma parameters. *Journal of Geophys-*

ical Research (Space Physics)*, **vol. 115**, A04106, 2010.
http://dx.doi.org/10.1029/2009JA014680.

[25] Roure D.D., Goble C., Stevens R.: The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows. *Future Generation Computer Systems*, **vol. 25**, pp. 561–567, 2009.

[26] Stone E. C., Frandsen A. M., Mewaldt R. A., Christian E. R., Margolies D., Ormes J. F., Snow F.: The Advanced Composition Explorer. *Space Science Reviews*, **vol. 86**, pp. 1–22, 1998. http://dx.doi.org/10.1023/A:1005082526237.

[27] Uchida Y.: *On the Exciters of Type II and Type III Solar Radio Bursts. Publications of the Astronimical Society of Japan*, **vol. 12**, p. 376, 1960.

[28] Webb D. F., Howard T. A.: Coronal Mass Ejections: Observations. *Living Reviews in Solar Physics*, **vol. 9**, p. 3, 2012.
http://dx.doi.org/10.12942/lrsp-2012-3.

[29] Yurchyshyn V., Yashiro S., Abramenko V., Wang H., Gopalswamy N.: Statistical Distributions of Speeds of Coronal Mass Ejections. *Astrophysical Journal*, **vol. 619**, pp. 599–603, 2005. http://dx.doi.org/10.1086/426129.

Affiliations

Gabriele Pierantoni
https://www.scss.tcd.ie/Gabriele.Pierantoni/, e-mail: gabriele.pierantoni@cs.tcd.ie

Eoin Carley
http://www.tcd.ie/Physics/Astrophysics/carley.php, e-mail: ecarley@tcd.ie

Jason Byrne
http://www.ifa.hawaii.edu/jbyrne/, e-mail: jbyrne@IfA.Hawaii.Edu

David Perez-Suarez
e-mail: dperez-suarez@sansa.org.za

Peter T. Gallagher
http://www.tcd.ie/Physics/people/Peter.Gallagher/, e-mail: peter.gallagher@tcd.ie