Gabriele Pierantoni
Brian Coghlan
Eamonn Kenny
Peter Gallagher
David Pérez-Suárez

EXTENDING THE SHEBA PROPAGATIO MODEL TO REDUCE PARAMETER-RELATED UNCERTAINTIES

Abstract

Heliophysics is the branch of physics that investigates the interactions and correlation of different events across the Solar System. The mathematical models that describe and predict how physical events move across the solar system (ie. Propagation Models) are of great relevance. These models depend on parameters that users must set, hence the ability to correctly set the values is key to reliable simulations. Traditionally, parameter values can be inferred from data either at the source (the Sun) or arrival point (the target) or can be extrapolated from common knowledge of the event under investigation. Another way of setting parameters for Propagation Models is proposed here: instead of guessing a priori parameters from scientific data or common knowledge, the model is executed as a parameter-sweep job and selects a posteriori the parameters that yield results most compatible with the event data. In either case (a priori and a posteriori), the correct use of Propagation Models requires information to either select the parameters, validate the results, or both. In order to do so, it is necessary to access sources of information. For this task, the HELIO project proves very effective as it offers the most comprehensive integrated information system in this domain and provides access and coordination to services to mine and analyze data. HELIO also provides a Propagation Model called SHEBA, the extension of which is currently being developed within the SCI-BUS project (a coordinated effort for the development of a framework capable of offering to science gateways seamless access to major computing and data infrastructures).

1. Introduction

In Heliophysics, relations among different events originated at the Sun are investigated with *Propagation Models* that describe and predict the movement of physical events through the Solar System. The available measurements allow heliophysicists to classify such events into three categories: *Coronal Mass Ejections* (CMEs), *Solar Energetic Particles* (SEPs) and *Corotating Interactive Regions* (CIRs). The first two are usually associated with explosions (*i.e.*, solar flares) that happen on the surface of the Sun, whereas the last is produced by open magnetic fields generated by the Sun.

The HELIO project [4, 7, 8, 14] has developed SHEBA [13], a Propagation Model of the ballistic kind. Depending on whether it is run from the Sun to the planets or vice-versa and on which phenomena is investigated, SHEBA requires information on a variety of parameters such as the position on the surface on the Sun from where the event originates, its time and its expected speed at the source. These parameters are not always known by the scientists; some can be inferred from observations of the Sun, others can be extracted from catalogues of events that are publicly available based on multiple observations or automated algorithms while others can be simply determined by experience.

This paper proposes a different approach: to run the model with varying parameters as a *parameter sweep job* and then select the optimal ones from the results.

To address this scenario, the concept of an *Advanced Propagation Model* has been introduced. Such a model runs a large number of models with different parameters and validates the results by looking into the event lists at the Sun and different planets.

From a computational point of view, propagation models vary greatly depending on their complexity; the SHEBA Propagation Model is relatively lightweight and allows for almost real time execution. Of course, once SHEBA is used within the Advance Propagation Model, then its execution time is determined by the selected range and granularity of the values.

The SHEBA propagation model was executed in the *HELIO Processing Service* (HPS) [14] described in section 4. This processing service executes jobs on two different resources depending on their execution time: short jobs are executed on a *Fast Execution Resource* that has very little overhead while longer jobs are executed on a gLite 3.2 [1] grid. The processing and storage services of HELIO served the needs of the project but lacked flexibility in adding and removing different computational resources as well as a poor connection with the user interface.

The SCI-BUS [6] project is expected to overcome these issues with its capabilities of handling multiple computational back-ends and its multi-layered architecture that offers functionalities and related APIs at each level, from the GUI to the resources.

This paper is structured as follows:

- *Propagation Models in Heliophysics*, in Section 2, introduces the use of Propagation Models in Heliophysics.

- *The Advanced Propagation Model*, in Section 3, introduces the Advanced Propagation Model.
- *The HELIO Project*, in Section 4, details the HELIO project with a special focus on the services used by the Advanced Propagation Model.
- *The SCI-BUS Project*, in Section 5, introduces the SCI-BUS project.
- *Architecture of the Advanced Propagation Model Portal*, in Section 6, details the architecture of the system.
- *Application fields of the advanced propagation model*, in Section 7, introduces various possible applications of the Advanced Propagation Model in Heliophysics.
- *Conclusions and Future Work*, in Section 8, concludes the paper and introduces future directions of research.

2. Propagation Models in Heliophysics

The heliosphere is a very dynamic environment: the planets and other bodies (*e.g.*, dwarf-planets, asteroids, spacecraft) move under the laws of gravitation, and the Sun and some of the planets generate magnetic fields that interact among each other. Finally, the solar activity and the solar wind shapes the magnetic field of the whole solar system.

The rotation of the Sun and the properties of the inter-planetary plasma shape the magnetic field of the heliosphere as a spiral [12]. This is key to understanding how events propagate, for example, high energy particles travel along magnetic field lines, but the regularity of the spiral shape can be disturbed by solar events like *coronal mass ejections* (CMEs).

Understanding how events propagate and interact in the heliosphere is fundamental to Heliophysics. To achieve this goal, different Propagation Models have been proposed over the years.

Before describing the different available models, it is essential to understand which instruments provide information that can be used to compare the results of the models. The list of available data can be separated into two groups: *remote* and *in-situ* observations.

The *remote* instruments include devices that take images of the solar surface and atmosphere, magnetographs that measure the surface magnetic field, and coronographs that observe the faint surroundings of the solar disk.

In-situ instruments are those that measure the properties of the solar wind and its environment; for example the strength of the local magnetic field, the velocity, temperature and density of particles.

Most of these observations are being processed almost in real time by manual or automatic algorithms and the resulting meta-data is stored in catalogues.

Although an increasing number of these instruments are on-board different spacecraft at different points in the heliosphere, we are still far from measuring the current state on the whole heliosphere.

Propagation models are used to infer information for the gaps that results from the uneven coverage of the solar system.

Some heliospheric models are based on magnetohydrodynamic (MHD) simulations produced from solar surface magnetic field observations[15]. These simulations can be adapted to include events like CMEs and to study the disturbances produced in the *quiet* solar wind. Most of these simulations are time and computational intensive, and they require knowledge of certain variables that are, at times, completely unknown (*e.g.*, CME's mass). Some of these models are used also to propagate and forecast the arrival of other events like *Corotating Interactive Regions* and *Solar Energetic Particles* (produced, for example, by solar flares).

Within the HELIO project (introduced in section 4), the SolarHeliospheric Event Ballistic Algorithm (SHEBA)[1] was developed. SHEBA is not data driven and it uses a simple ballistic model, *i.e.*, all the propagation is done assuming a constant speed, though this approximation could seem very inaccurate, it has been shown that it is not far from reality during quiet periods [13]. The main aim in SHEBA was to offer a fast and easy propagation model to the users of HELIO. The scientists that have seen an event on the Sun or found one event of interest in any of the events catalogues available, will use SHEBA to understand when and where to look for the signature of this event in other places of the heliosphere.

SHEBA provides within seconds the expected time ranges of arrivals (ETA) for the different locations in the heliosphere that are under investigation.

Though the event catalogues mentioned above could provide some valuable information to the parameters of the propagation model, it is still not possible to infer all the initial parameters from existing data sets. For example, the remote observations of CMEs cannot reliably provide an estimate of the speed at which the event leaves the surface of the Sun. Only an estimate of the lower limit of speed can be deduced by the movement of the CME on a 2D projection. Setting reliable values of other parameters, such as the longitudinal width, proves to be equally complicated.

3. The Advanced Propagation Model

3.1. Propagation Models

As we introduced in sections 1 and 2, a *Propagation Model* in Heliophysics allows us to simulate and predict the movement of events across the solar system.

A complete definition of Propagation Models is outside the scope of this paper; for the sake of this investigation, we can simplify a Propagation Model as a mathematical function that predicts the arrival time at a target of a physical event that departed from another position in space (*origin*).

$$\widehat{t_t} = F_{pm}(t_o, \overrightarrow{x_o}, \overrightarrow{p}, \overrightarrow{x_t}) \tag{1}$$

---

[1]SHEBA is open-sourced and the code is accessible: `https://github.com/dpshelio/sheba` and the service is set at `http://cagnode58.cs.tcd.ie/PropagationModelGUI`.

where:

- $\widehat{t_t}$ is the expected arrival time on the target,
- F_{pm} is the propagation model function,
- t_o is the time when the physical event left the origin,
- $\overrightarrow{x_0}$ is the position of the origin,
- $\overrightarrow{p}$ are the additional parameters for the model, these usually contain at least an estimation of the speed of the phenomenon at the origin,
- $\overrightarrow{x_t}$ is the position of the target.

It is worth mentioning that Propagation Models are *reversible* in the sense that they can predict the departure time of a physical features given an estimate of the speed and the arrival time at the target. In this case, the Propagation Model is:

$$\widehat{t_o} = F_{pm}(t_t, \overrightarrow{x_t}, \overrightarrow{p}, \overrightarrow{x_o}) \tag{2}$$

Also, many Propagation Models give an *estimated range of times* instead of just an expected arrival time; in this case, the Propagation Model is:

$$\widehat{\Delta_{t_t}} = F_{pm}(t_o, \overrightarrow{x_o}, \overrightarrow{p}, \overrightarrow{x_t}) \tag{3}$$

where:

$$\widehat{\Delta_{t_t}} = \left[\widehat{t_{t_{min}}}, \widehat{t_{t_{max}}}\right] \tag{4}$$

When propagation models return a range of possible expected arrival times $(\widehat{\Delta_{t_t}})$ instead of just an expected arrival time $(\widehat{t_t})$, they require at least the estimated speed of the event at source and an associated estimated error in the form $\widehat{s} \pm \widehat{e}$

Advanced Propagation Models do not just apply their mathematical model but also validate their results by querying catalogues of events. For this reason we introduce *Ancillary functions* such as *Querying Functions* and *Ranking Functions*

3.2. Querying functions

Querying functions are used to find events within a given time range, they are formalizations of the execution of queries over catalogues of events.

$$\overrightarrow{ev} = F_q\left(t_{min}, t_{max}, \overrightarrow{ct}\right) \tag{5}$$

where:

- $\overrightarrow{ev}$ is the sequence of events returned by the query function,
- F_q is the querying function,
- t_{min} and t_{max} are the boundaries of the time range,
- $\overrightarrow{ct}$ is the list of catalogues of events used,

and where each *event* is modelled as

$$ev_i = \left\{t_e, \overrightarrow{p_e}\right\} \tag{6}$$

where:

- t_e is the time of the event. It is worth mentioning that, depending on the catalogue used, there can be different types of times associated with an event such as *start*, *end* or *peak time*.
- $\overrightarrow{p_e}$ are additional parameters of the event.

3.3. Ranking functions

Ranking functions are used to rank events depending on different parameters.

$$r_i = F_r(ev_i) \tag{7}$$

where:

- r_i is the rank associated by the ranking function to the event,
- F_r is the ranking function,
- ev_i is the event being ranked.

3.4. The SHEBA Propagation Model

The advanced propagation model presented here is based on SHEBA (introduced in section 2). Though SHEBA has very simple inputs, some of them cannot be known from the current observations available (*i.e.*, velocity of the CME and its longitudinal width), and it requires considerable experience to choose reasonable values.

The parameters that define the propagation model depend on the event under investigation and the direction of the simulation (from the Sun to a target or vice-versa). The following images and tables list the parameters and their meaning.

The relationship between the parameters and the propagation model of Coronal Mass Ejections (CME) is illustrated in Figure 1; their meaning for the forward model (from the Sun to the target) and the backward model (from target to the Sun) are detailed in Table 1 and 2. Likewise for Corotating Interactive Regions see Figure 2 and Table 3 and 4, and for Solar Energetic Particles (SEP) see Figure 3, and Table 5 and 6.

In the advanced propagation model, SHEBA can be used in two ways: for a single event and for multiple events. At present the Advanced Propagation Model only includes the SHEBA CME forward propagation functionality.
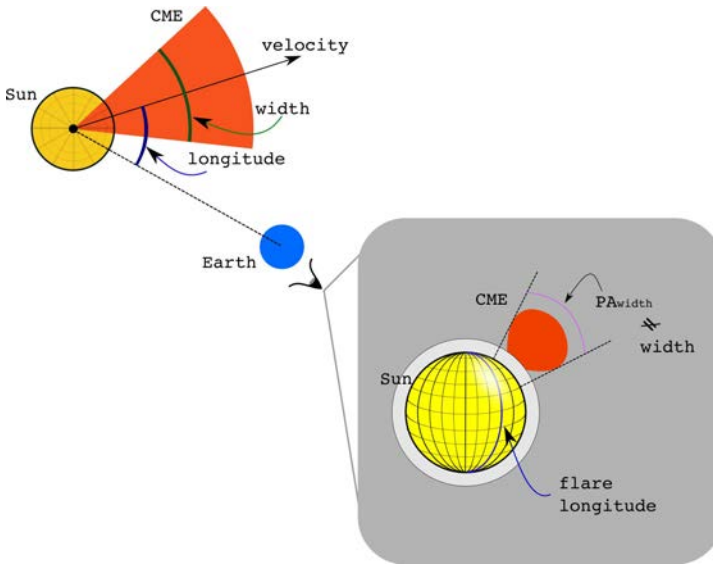
3.5. Single event

The Advanced Propagation Model can be executed on a single event to determine parameters in the model that cannot be easily defined by the user. Such a scenario is illustrated in Figure 4.

The first step is to execute the propagation model for one particular event and a range of parameters.

$$\widehat{\Delta_{t_{t_i}}} = F_{pm}(\widehat{t_o}, \overrightarrow{x_o}, \overrightarrow{p_i}, \overrightarrow{x_t}) \tag{8}$$

with

$$p_{min} < p_i < p_{max} \tag{9}$$

Figure 1. The different parameters that govern the propagation of a Coronal Mass Ejection in the SHEBA Propagation Model.
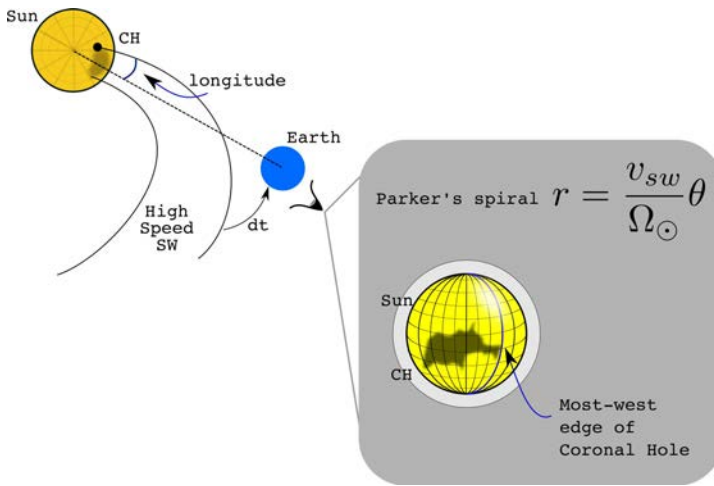
Table 1

Parameters that govern the forward simulation of CME events in the SHEBA Propagation Model.

Parameter	Meaning
Start Time	The time at which the event departed the Sun
Longitude	The relative angle between the starting poing of the event and the Earth
Width	The width of the CME
Speed	The speed at which the CME left the surface of the Sun
Speed Error	The error in the speed

Table 2

Parameters that govern the backward simulation of CME events in the SHEBA Propagation Model.

Parameter	Meaning
Start Time	The time at which the event departed the Sun
Object	The target from which to backtrack the event to the Sun
Width	The width of the CME
Speed	The speed at which the CME left the Sun
Speed Error	The error in the speed

Figure 2. The different parameters that govern the propagation of a Corotating Interactive Region in the SHEBA Propagation Model.

Table 3

Parameters that govern the forward simulation of CIR events in the SHEBA Propagation Model.
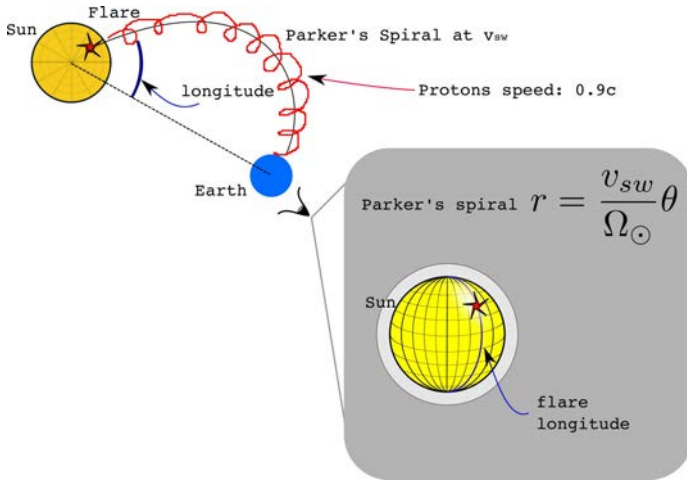
Parameter	Meaning
Start Time	The time at which the event departed the Sun
Longitude	The relative angle between the starting poing of the event and the Earth
Speed	The speed at which the CIR left the Sun
Speed Error	The error in the speed

Table 4

Parameters that govern the backward simulation of CIR events in the SHEBA Propagation Model.

Parameter	Meaning
Start Time	The time at which the event departed the Sun
Object	The target from which to backtrack the event to the Sun
Speed	The speed at which the CIR left the Sun
Speed Error	The error in the speed

Figure 3. The different parameters that govern the propagation of a Solar Energetic Particles in the SHEBA Propagation Model.
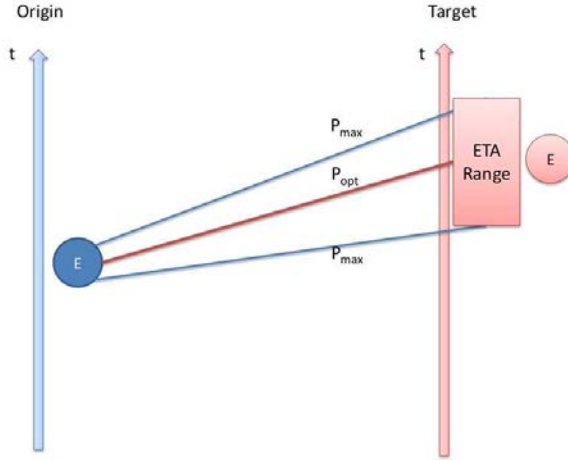
Table 5

Parameters that govern the forward simulation of SEP events in the SHEBA Propagation Model.

Parameter	Meaning
Start Time	The time at which the event departed the Sun
Longitude	The relative angle between the starting poing of the event and Earth
Speed	The speed at which the SEP left the surface of the Sun
Speed Error	The error in the speed
Beta	The travelling speed of energetic particles as percent of speed of light

Table 6

Parameters that govern the backward simulation of SEP events in the SHEBA Propagation Model.

Parameter	Meaning
Start Time	The time at which the event departed the Sun
Object	The target from which to backtrack the event to the Sun
Speed	The speed at which the SEP left the surface of the Sun
Speed Error	The error in the speed
Beta	The travelling speed of energetic particles as percent of speed of light

Figure 4. The outcome of the execution of the Advanced Propagation Model on a single event. The variation of a single parameter P between P_{min} and P_{max} results in the range $ETARange$ for the Expected Arrival Times (ETA). The optimal parameter P_{opt} is found in correspondence of the event E at the target

The second step is to use the query functions to find relevant events at the target within the expected time of arrival.

$$\overrightarrow{ev_i} = F_q\left(\widehat{\Delta_{t_{t_i}}}, \overrightarrow{ct}\right) \tag{10}$$

By composing equation (8) and (10), we obtain:

$$\overrightarrow{ev_i} = F_q\left(F_{pm}(\widehat{t_o}, \overrightarrow{x_o}, \overrightarrow{p_i}, \overrightarrow{x_t}), \overrightarrow{ct}\right) \tag{11}$$

The third step is to use a ranking functions on the event.

$$r_i = F_r(ev_i) \tag{12}$$

By composing equation (11) and (12), we obtain:

$$r_i = F_r\left(F_q\big(F_{pm}(\widehat{t_o}, \overrightarrow{x_o}, \overrightarrow{p_i}, \overrightarrow{x_t}), \overrightarrow{ct}\big)\right) \tag{13}$$

Finally, we select the parameters (p_{opt}) that correspond to the highest rank.

At the moment, the ranking function is a binary function that returns one if an event is found at the target and zero otherwise. The parameter is set to be optimal when the ranking function associated to it evaluates to 1.
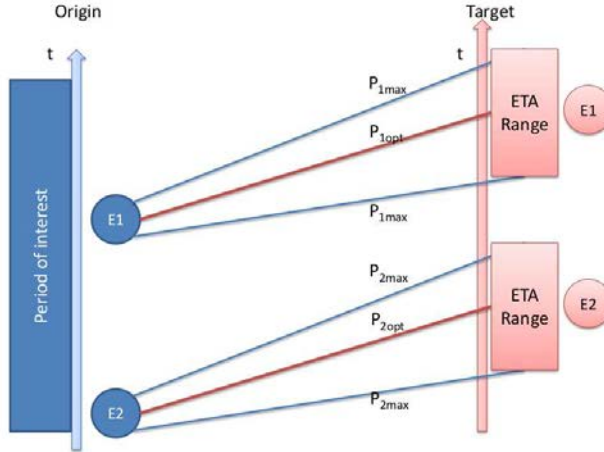
$$p_{opt}: \max_{p_{min} \leq p_i \leq p_{max}}\left(F_r\big(F_q(F_{pm}(\widehat{t_o}, \overrightarrow{x_o}, \overrightarrow{p_i}, \overrightarrow{x_t}), \overrightarrow{ct})\big)\right) \tag{14}$$

So, the optimal execution of the Propagation Model becomes:

$$\widehat{\Delta_{t_t}} = F_{pm}(\widehat{t_o}, \overrightarrow{x_o}, \overrightarrow{p_{opt}}, \overrightarrow{x_t}) \tag{15}$$

3.6. Multiple events

Another use for the Advanced Propagation model is to execute the model for all events in a given time range. This use of the model is described in Figure 5.

Figure 5. The outcome of the execution of the Advanced Propagation Model on multiple events. The variation of a single parameter for each event P_i between P_{imin} and P_{imax} results in the range *ETA Range* for the Expected Arrival Times (ETA) for each event. The optimal parameters P_{iopt} for each event are found in correspondence of the events E_i at the target.

First, we use the Querying function to obtain all the events at the source that fit into a time range.

$$\overrightarrow{ev} = F_q(t_{min}, t_{max}, \overrightarrow{ct}) \tag{16}$$

Then, we apply equation (14) to each event to find the optimal parameters for each event and obtain the propagation model executions with optimal parameters for each event.

3.7. Catalogues

The current implementation of the Advanced Propagation Model relies on different catalogues to find events at the Sun and other targets. These are listed in Table 7 for the catalogues of events on the surface of the Sun, and in Table 8 for the catalogues of events on Earth, and in Table 9 for the catalogues of events on Mars.

4. The HELIO project

HELIO, a Research Infrastructures funded under the Capacities Specific Programme within the European Commission's Seventh Framework Programme (FP7; Project

Table 7

Event Catalogues used by the Advanced Propagation Model to spot events at the Sun.

Sun
GOES Soft X-ray Flare List
NGDC H-alpha Flare List
RHESSI Hard X-ray Flare List
Ulysses/GRB X-ray Flare List
Yohkoh/HXT Hard X-ray Flare List
SOHO/LASCO CME Event List
CACTus SOHO/LASCO CME List

Table 8

Event Catalogues used by the Advanced Propagation Model to spot events on Earth.

Earth
SOHO/CELIAS/MTOF/PM Interplanetary Shock List
NGDC AA* Major Magnetic Storm
NGDC Ap* Major Magnetic Storm
CME-related Forbush Decrease Event List
AAD Ground Level Enhancement List

Table 9

Event Catalogues used by the Advanced Propagation Model to spot events at Mars.

Mars
Mars Earth ICME

No. 238969), provides a comprehensive information system for heliophysics. The project started on 1 June 2009 and will be completed by 30 November 2012.

To be an effective research infrastructure, HELIO offers support across all the different steps that a scientist usually performs in his inquiry. A common, albeit not entirely comprehensive, approach to scientific inquiry in Heliophysics can be decomposed into four main steps.

- The first step (*metadata search*) is devoted to identifying interesting phenomena. During this phase, the user interrogates specific HELIO services that query catalogues of events and features and uses only metadata which has been extracted from previous observations.

- The second step (*instrument search*) consists in the search of instruments that have observed the events and features of step 1. During this phase, the user uses a service that queries the positions of instruments scattered across the solar system, a propagation model to track the movement of phenomena from the sun

to the outer regions of the solar system, and a Coordinate Transformation Service (CTS).

- The third step (*observations search*) uses the list of instruments obtained in the second step to review the availability of suitable observations by querying lists with the position and type of observations of every instrument. During this phase, the user uses HELIO services that query catalogues of instruments by capabilities, position and status at any given time.

- The fourth step (*data search*) consists in locating, selecting and eventually retrieving the data related to the observations selected in step 3. During this phase, the user uses a service, called DPAS (Data Provider Access Services) that knows which types of data are held at what place and how to access this data.

This approach is widely used but does not cover entirely all the possible approaches in the Heliophysics community; to cater for all this variety, HELIO has been designed with a multi-layered architecture (described in Figure 6) that offers many different routes to each of its services both individually and jointly.
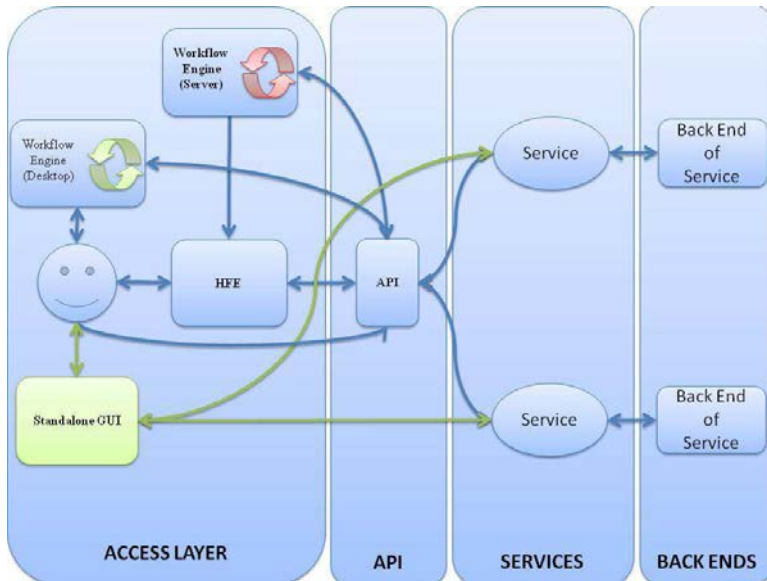
Figure 6. An overview of the architecture implemented in the HELIO project that illustrates its different layers and their purpose.

Most HELIO services expose their own standalone graphical user interface but they are also all connected to a centralized graphical user interface named HELIO Front End (HFE), a browser-based user interface intended to solve the most common tasks.

The HFE is also the best choice for less experienced users, both for its user-friendliness and for its capability to allow users to connect services to each other

to explore new orchestration patterns. For more advanced users HELIO exposes an IDL [5] (Interactive Data Language) API that connects the system directly to custom-made IDL programs, and for expert users HELIO offers programmatic access through a Java library (HELIO Client API) that simplifies this access by providing client stubs for services. Furthermore, HELIO services can also be handled through the Taverna Desktop workflow tool [11] to orchestrate several services into more complex workflows.

Finally, HELIO not only accesses existing metadata information sources or observational data but also allows data processing to create new metadata catalogues or derived results through its specific services that offer computation, storage and an IDL environment to execute metadata extraction algorithms.

Of all the services that HELIO offers, three are most important for this research: The HELIO Event Catalogue, The SHEBA Propagation Model and the HELIO Processing Service.

The HELIO Event Catalogue (HEC) is a service that can be queried for events such as CME's.

The SHEBA propagation model has been described in Section 3.

The HELIO Processing Service (HPS) is an interface to different computational resources that offers to HELIO two different kinds of services: the execution of a pre-defined set of applications and the execution of user-defined applications. The HPS supports two different computational resources. A fast execution engine suitable for short lived jobs (where SHEBA runs) and a more powerful (albeit with a longer overhead time) computational resource based on a grid, i.e. via a gLite workload management system (WMS). The architecture of the HPS is described in Figure 7.
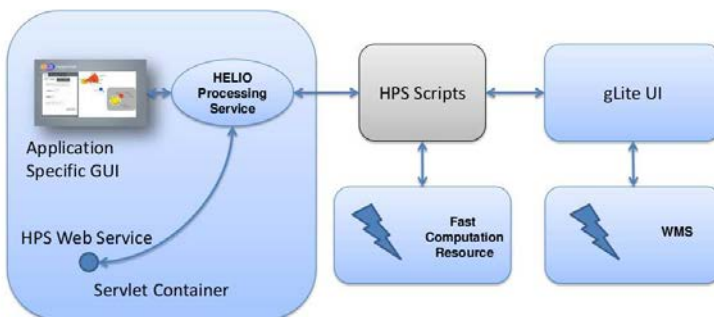

Figure 7. A detail of the HELIO architecture that shows the different components of the HELIO Processing Services (HPS).

5. The SCI-BUS project

SCI-BUS is an FP7 effort that aims to create a generic-purpose gateway technology that provides access to clusters, supercomputers, grids, desktop grids, academic and commercial clouds.

SCI-BUS is based on the WS-PGrade/gUSE technology [10] and its aim is to support a large number of different communities in building portals, implementing execution logic based on workflows and in accessing diverse *Distributed Computing Infrastructures*(DCI).

The SCI-BUS architecture is described in Figure 8; it comprises several layers each capable of offering the ideal level of complexity to the portal developer.
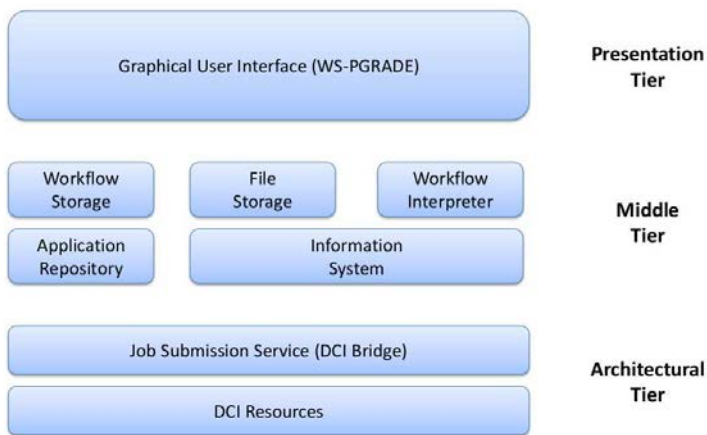
Figure 8. An overview of the different layers and components of the SCI-BUS architecture.

The top layer, based on WS-PGRADE is known as the *Presentation tier* and offers a ready-to-use graphical user interface to those who do not want to develop their own custom Grapgical User Interface. This layer offers powerful functionalities to design and execute workflows on heterogenous Distributed Computing Infrastructures. The Graphical User Interface also offers functionalities for file management, storage for workflows and applications, monitoring and management security, with special attention to grid certificates.

For users and communities who need a different interface and have the resources to develop their own portal, SCI-BUS offers a rich, multi-layered set of APIs and services that offer at a programmatic level the same functionalities that are exposed by the top layer but allow for a complete customization to complete the needs of the users.

The middle layer (*Middle tier*), offers a set of APIs and services for the management of the most common functionalities that are required by an e-science portal such

as the interpretation and storage of workflows, storage and execution of applications, file management and the querying of DCI information systems.

The bottom layer (*Architectural tier*), offers an API for job submission on a large variety of Distributed Computing Infrastructures, which is of primary interest to all those portals that focus on the execution of applications rather than the services available on the Middle tier.

6. Architecture of the Advanced Propagation Model Portal

The functions described in Section 3 are implemented by different modules written in Java:

- *The SHEBA propagation model* is implemented as a set of IDL routines that are executed via scripts in one of the computational resources shown in Figure 9.
- *The querying functions* are implemented as a set of invocations to the HELIO Event Catalogue (HEC).
- *The ranking functions* are implemented directly in the controller(see Figure 9).
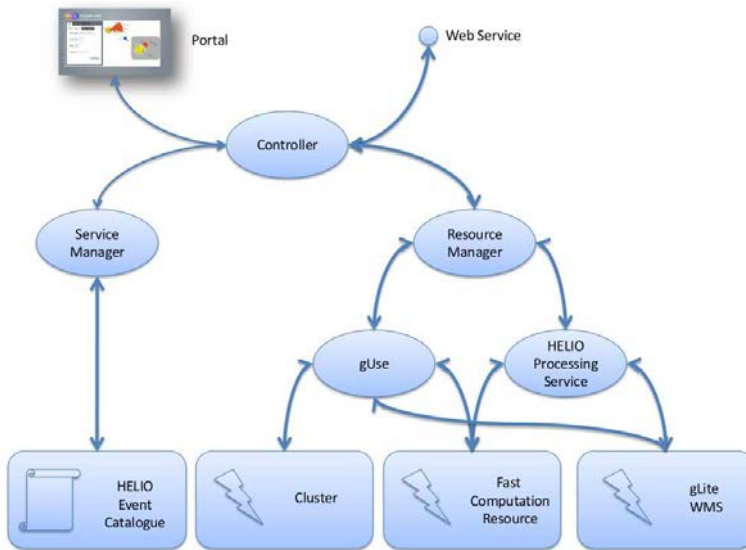
Figure 9. An overview of the different components of the Propagation Model Portal.

The architecture of the first prototype is sketched in Figure 9. The advanced propagation model exposes two interfaces: a portal hosted in a Liferay environment and a Web-Service Interface.

Both interfaces are connected to the controller that uses the resource manager and a service manager. The controller hosts the logic of the service; it determines the number of executions of the propagation model from the event lists obtained

from the event catalogues, it controls the execution of the propagation model and it defines the ranking of the events used to validate the results in the last steps. The service manager wraps the client to the legacy HELIO services such as the HEC (HELIO Event Catalogue) and it extracts suitable information from the VOTable [9] returned; this abstraction layer is particularly useful as it will allow expanding the system to more services (possibly also from projects other than HELIO such as the HEK [3]).

The resource manager acts as an abstraction layer to the processing service inherited from HELIO and extends it with the services offered by SCI-BUS. There are at present two computational resources: the Grid-Ireland gLite grid and a fast computational resource in Trinity College Dublin that executes light jobs without the extra time required by the gLite middleware. These two middlewares are connected through the HELIO Processing Service (HPS) but by the end of 2012, Grid Ireland will decommission its grid middleware (the UK GridPP will take over grid support for HELIO) and Trinity College Dublin will partly devote a cluster managed by PBS to the Advanced Propagation model portal, while other clusters will be available as part of academic clouds. The flexibility that SCI-BUS offers in accessing DCI resources is a perfect fit for both the needs of the Advanced Propagation Model (fast computation resource for a single execution of the model, cluster and grid resources for the parameter sweep job) and the changing circumstances at Grid-Ireland.

7. Application fields of the Advanced Propagation Model

At the moment the Advanced Propagation Model proposed in this paper is based on the propagation of CMEs and depends on the existence of catalogues within HELIO to find the parameters that best match. Within HELIO this could be easily adapted to the case of Corotating Interactive Regions and Solar Energetic Particles events, by just changing the input parameters and the catalogues to cross-check. However, there are other applications for which this method could also be used. Firstly, a large run over the whole space-observation age could provide valuable statistics on the accuracy of the proposed model in a reasonable amount of time. These statistics could help for further development on the underlying propagation model (*e.g.*, the addition of a dragging parameter/function which decelerates the CMEs). These statistics could also be used to develop a forecasting system, where the CMEs observed in real-time could be compared with previous ones and automatically set the parameter ranges obtained from the most similar ones. Finally, the results provided from the Advanced Propagation Model could be used as boundary limits for larger computational models (*e.g.*, that based on MHD simulations).

At the moment, the advanced propagation model depends on catalogues and this limits its scope of action to the study of events that were manually or automatically detected. A way to overcome this limitation is to connect the advanced propagation model to the automated detection codes of in-situ phenomena In this case, the advanced propagation model would not need to access catalogues to find new corre-

lations. This feature would also have the benefit of linking multiple measurements at different places in the heliosphere that have been not catalogued yet, and could enable automated insertion of newly detected events into catalogues.

8. Conclusions and future work

Although the Advanced Propagation Model Portal is currently in a prototypical stage, the approach has proven useful to define a subset of parameters for a more accurate execution of the model. The greatest advantage of using the Advanced Propagation Model has proven to be not only a less arbitrary setting of the value of the parameters but also a closer relation between the model itself and the event catalogues used to validate the results.

The exploitation of the full potential of the system is yet to be fully discovered and and close interaction with the Heliophysics community is planned to ensure maximum usability for the community.

The SCI-BUS project has planned associated periods of development and gathering of users feedback to allow an effective cooperation among the developers and the scientific communities.

From a technological point of view, closer integration with the SCI-BUS functionalities of the *middle tier* will be required to use workflows to enable better flexibility in the definition and implementation of the execution strategies of the parameter-sweep jobs.

Regarding functionalities, the most urgent task will be to include the models for *High speed solar wind* and *Solar energetic particle events* in addition to the *Coronal Mass Ejection* currently supported.

Finally, to foster the usability of the portal, we plan to build a database with all the results of past executions of the model, to allow the search of events to be ranked according to the quality of matching. This will open an opportunity to provide a semantic search engine, where the user could put certain keywords (e.g. ”Flare Jupiter”) to obtain from the system all matches containing a flare as a precursor and Jupiter as an observational point. This list of results should be based on the quality factor calculated by the Advanced Propagation Model.

Acknowledgements

References

[1] glite web page – `http://glite.cern.ch/`.

[2] Grid ireland web page – `https://grid.ie/`.

[3] Hek project page – `http://www.lmsal.com/hek/api.html`.

[4] Helio project page – `http://www.helio-vo.eu/`.

[5] Idl project page – `http://www.exelisvis.com/language/en-us/productsservices/idl.aspx`.

[6] Sci-bus project page - https://www.sci-bus.eu/web/guest.

[7] Bentley R. D.: Building a virtual observatory for heliophysics. *Earth, Moon, and Planets*, 104:87–91, 2009.

[8] Bentley R. D., Csillaghy A., Aboudarham J., Jacquey C., Hapgood M. A., Bocchialini K., Messerotti M., Brooke J., Gallagher P., Fox P., Hurlburt N., Roberts D. A., Duarte L. S.: Helio: The heliophysics integrated observatory. *Advances in Space Research*, 2009.

[9] Franois O., Roy W., Clive D., Daniel D., Pierre F., Robert H., David G., Tom M., Alex S., Andreas W.: Votable: Tabular data for the virtual observatory. In *Toward an International Virtual Observatory*, Springer-Verlag, p. 118, 2004.

[10] Kacsuk P., Farkas Z., Kozlovszky M., Hermann G., Balasko A., Karoczkai K., Marton I.: WS-PGRADE/gUSE Generic DCI Gateway Framework for a Large Variety of User Communities. *Journal of Grid Computing*, pp. 1–30, November 2012.

[11] Oinn T., Greenwood M., Addis M., Alpdemir M. N., Ferris J., Glover K., Goble C., Goderis A., Hull D., Marvin D., Li P., Lord P., Pocock M. R., Senger M., Stevens R., Wipat A., Wroe C.: Taverna: lessons in creating a workflow environment for the life sciences. *Concurrency and Computation: Practice and Experience*, 18(10):1067–1100, 2006.

[12] Parker E. E. N.: Dynamics of the Interplanetary Gas and Magnetic Fields. *The Astrophysical Journal*, 128:664, November 1958.

[13] Pérez-Suárez D., Maloney S. a., Higgins P. a., Bloomfield D. S., Gallagher P. T., Pierantoni G., Bonnin X., Cecconi B., Alberti V., Bocchialini K., Dierckxsens M., Opitz A., Blanc A., Aboudarham J., Bentley R. B., Brooke J., Coghlan B., Csillaghy A., Jacquey C., Lavraud B., Messerotti M.: Studying SunPlanet Connections Using the Heliophysics Integrated Observatory (HELIO). *Solar Physics*, 280(2):603–621, September 2012.

[14] Pierantoni G., Coghlan B., Kenny E.: The architecture of helio. In *Krakow Grid Workshop 2010*, 2010.

[15] Xie H., Ofman L., Lawrence G.: Cone model for halo CMEs: Application to space weather forecasting. *Journal of Geophysical Research (Space Physics)*, 109:3109, March 2004.

Affiliations

Gabriele Pierantoni

Trinity College Dublin, TCD Dublin 2, `pierang@cs.tcd.ie`

Brian Coghlan
    Trinity College Dublin, TCD Dublin 2, `coghlan@cs.tcd.ie`

Eamonn Kenny
    Trinity College Dublin, TCD Dublin 2, `ekenny@cs.tcd.ie`

Peter Gallagher
    Trinity College Dublin, TCD Dublin 2, `peter@email`

David Pérez-Suárez
    Finnish Meteorological Institute, Helsinki, Finland, `david.perez-suarez@fmi.fi`