Scientific Challenges Pertaining to Space Weather Forecasting Including Extremes

Pasadena, CA, USA • 11-15 February 2019 CHAPMAN CONFERENCE

Discussion Notes

Tuesday, February 12, 2019

Scribes: Anthony Mannucci, Xing Meng, Anna Kelbert. Assistance from Delores Knipp

Note: To encourage free communication during and after the meeting comments are not attributed.

This document reports on major Discussion Topics (Themes) on Day 1 of the Chapman Conference Scientific Challenges Pertaining to Space Weather Forecasting Including Extremes. These themes were derived from the pre- and early-meeting survey results.

1) **Temporal and spatial scales for practical forecasts**
2) **Metrics**
3) **Forecast lead times**
4) **Forecast analogies**
5) **Users**
6) **Observations**
7) **Extremes**
8) **Uncategorized**

Discussion Topic that started the discussion (based on survey results):
It is important to define temporal and spatial scales for which forecasting is practical

Theme: Temporal and spatial scales for practical forecasts

I want to understand what the spatial/temporal scales mean.

I take it to mean regional vs local, spatially. Temporally, different time scales involved in different parts of the process.

You can talk about flares, CMEs, propagation time. All of that is physics. But temporal scale may mean the time that the user has to respond to a warning.

How much time do we need to run the modeling? 3) How soon is the result needed? 4) What time scale to we use for modeling?

These drive our time and spatial scales.

Temporal and spatial scales are highlights of the survey, community focus.

What are the temporal and spatial scales people need, different answers for data vs models.

Forecast window, how accurate it needs to be: a huge gap between research and users. What is needed is a metric defined by users.

Current documents are too general. Need lots more communications with users. CCMC's done great job, model adjustments/metrics for different users/applications.

This group is not tasked with adding to that general work towards end user needs. The idea is to be transformative. Not what the end user needs, but what we can actually achieve. There are certain things we can't achieve, and I'm not sure it's well defined at this point.

Predicting from the beginning is good but not for every second after. Predicting Bz is chaos, practical limitations from physics and models and missing data.

Weather forecasts have limitations and people know that.

Agreed. E.g., SWPC challenge to predict magnetic perturbation at Earth's level – it was clear from the beginning that second by second, localized prediction is impossible. There are practical limitations from missing physics, model uncertainty etc.

You still need to know what happens to the CME while it travels through interplanetary medium.

But we know that is not possible, so the requirements evolved over time.

Do we have at least some indication of what kind of precision in terms of time/spatial scales is even possible? E.g., hourly means within a 1000 km square at the Earth's surface? No answer to that.

I think that's an extremely difficult problem, as far as understanding what's deficient and why it's deficient. That the models will improve over time is a widespread idea, but it's not obvious. E.g., hurricane tracks have improved but the intensity hasn't improved over the last 20 yrs. Let's run these models and understand what we have, that's what the metrics are for. But what information do we need to be confident that the metric will improve over time? What data will lead to improvements?

Even if Bz is known from CME images, still need to know how CME travels through the interplanetary medium, makes a big difference on geoeffectiveness of CMEs.

Temporal & spatial scales, do we know specific scales that forecasts can reach?

Forecast probabilities, i.e. during solar max things are more probable.

Turbulence in the magnetic cloud and plasma sheath (solar wind) make prediction very hard.

No need to predict CME? Depends on the phenomena, flares (minutes) vs GIC (days).

CMEs at sun will not all hit Earth, time scales play a role.

Do we have understanding of uncertainties? Not being addressed sufficiently.

Need to run models continuously in operational mode, see how they work for real events, feed back to model improvement. Extremely difficult problem (Tony).

Modelers can instruct on what data are needed for running the model in real-time. Data providers benefit from it.

Modelers are not being paid for developing models for operations. Things are changing… funding available for transitioning models to operational mode. For example, NASA's Space Weather Science Applications (SWxSA) program.

Physics models are needed for understanding, Final forecasts are not necessarily made by these models, perhaps ML.

Theme: Metrics

CCMC and other groups help standardize metrics and formats. E.g., the CME arrival scoreboard. Do we have new requirements and metrics that multiple groups can go and refer to?

Users don't care about model validation metrics.

We have all these models and we need to run them consistently in real time mode, even if they are not providing reliable results. Apply metrics to them periodically. This is a type of operations to research feedback.

Need bench-marking with possible range of outcomes.

What metrics to use in bench-marking whether models do a good job?

SWPC cannot do validation for every case/model. Validation has to be done by modelers/researchers.

Theme: Forecast lead times

We can do something similar to hurricane prediction for CME prediction. Spacecraft at L1/L5, cover the whole space from sun to earth. We don't know when an CME will occur, but once it occurred, it can be tracked continuously. Bz prediction: first improve the measurement before improving models, 2D mapping of magnetic structure of CMEs.

Theme: Forecast analogies

The weather forecasters stop a certain time in the future.
Discussion: Yes, you can theoretically prove that weather forecasts beyond a certain time (10 days?) are impossible. There are sufficient conditions for hurricanes but it doesn't mean that a hurricane will form. A 30-day forecast is impossible no matter how much money you throw at the problem. But it still might

be useful for users to know where hurricanes are likely to be, what their intensity is likely to be at what times in the year. They are not going to tell you 30 days ahead the trajectory of the hurricane. But they can say things like “This is an El Nino year, expect more rain”.

Discussion: You mentioned hurricane tracking. We can do something similar. We cannot predict where the CME will occur but once it’s occurred, we can follow the CME just like a hurricane.

In fact, what we’re going after is more akin to earthquake or volcano forecasting. Can’t predict until it actually happens.

I disagree with this view that it’s more like volcanoes. I like Jeff’s comment about probabilities. E.g., during solar maximum things are more improbable.

We do ourselves a disservice by comparing to hurricane prediction, where the phenomenon gradually forms. In our case, it’s a sudden eruption followed by turbulent behavior.
For hurricanes, they do ensemble modeling. Is that something that we can do with CMEs?
If we see a CME on the Sun, it doesn’t mean it hits the Earth. The thing can miss us. Even if it hits the Earth, we don’t know how it hits the Earth. But we still have minutes to hours to provide the useful prediction.

We need to start talking to other communities that do predictions of sorts.

Geophysicists are trying to understand the Earth, and eventually this could lead to forecasts. Why is an earthquake scientist not considered “operational”? What about a climate scientist? Then why are we (space weather community) considered operational?

30-day hurricane forecast is not going to happen in any case.

Useful to know when/where hurricanes are likely to be, there are dimensions that can be forecasted and useful to users.

Earthquake & volcano eruption prediction: many more measurements than for space weather, but still predictions are not possible. Are space weather prediction down to hour scale even possible?

Forecasting has a scientific aspect, not only operational.

Earthquakes/climate forecasts are not considered operational, why do we consider research on forecasting operational?

Hurricane prediction, most probable path, is based on current climatological models and ensemble modeling. Can we do something similar? Is it more beneficial for CME arrival at Earth? More probabilistic prediction of what may or may not happen for end users.

Theme: Users

1) What are the products needed for the users?

We have a wishlist from the users, but there's a huge gap between that and what's feasible. We have to bridge that somehow. Also, the metrics depend on what you need the data for.

A lot of the documents on what end users need are too general. Need something much more specific to inform the space weather forecaster/tools. Need more communication with the community on where we can find different models, what type of tools are there, and the needs.

And adopt their models to what the end user needs?
When you talk about detailed requirements by the end users, they won't tell you any real numbers because that will reveal proprietary information.
Example of successful collaboration with the industry: Antti Pulkkinen's effort on GICs.

Users want a forecast that's reliable, with uncertainty constrained.

The user wants more and more but we should pay the most attention to what's possible.
Perhaps we try to forecast even when we are not ready to do so. For example, when we expect Kp=9 or 8 and we succeed, but nothing happened, and the power company is misled. The user may not be worried about something like that. We need to work with the users.
Tony: Suppose that we work with the user and we learn what the user wants, but we can't deliver what the space weather [user] wants.

To go back to the point of working with users, satellite operators and designers. They would like forecasts up to a day ahead.

Also, what is happening now is very important for them. So, nowcast is also very important. If an event is taking place, how long is it going to last? When is it safe to resume operations? Are we actually forecasting the right sort of thing? For satellites, we have electron flux. But we try to look at engineering effects and tailor our forecasts more directly to the requirements.
Sounds like this evolution is a two-way street. A product of discussion.

Barrier on detailed requirements by users: users won't tell.

Try to forecast even if not ready, work with users like power grid company and figure out what they want.

Satellite operators need forecasts that are up to date, understanding nowcast is important. Are we forecasting the right things?

Prediction depends on user needs

What data do we need to be confident it will lead to model improvement?

No users 10-20 years ago, now we have users and know what they need. But it is moving target.

Theme: Observations

We want to have data everywhere but that's not practical.

Second, coronal magnetic field measurements. Isn't the coronal magnetic field measured at some level? – No. But they would be used to predict Bz after the eruption, or even before. But if we know Bz at the Sun, do we know it at L1?

It's also important that we start using the right data for the models that will be transitioned to operations. Taking research models and turning them into operational need appropriate data sets. Running models on bad/uncleaned data in real time isn't necessarily useful for end users.

Tony: in New Jersey they talked about that, better integration between the models and observations. Once we see what data is needed for us to run these models in real time, that gives us (and funding agencies) motivation to provide these data. So we can benefit from that.

Sometimes we use non-operational data, in order to test it really well before we use it in operations. We could do that with models.

Research models -> operational models need to deal with insufficient data/data not available in real-time. Need better integration between models and observations.

Theme: Extremes

We can also potentially be better at benchmarking. We need multiple realizations for benchmarks, using the same machinery as forecasting. What are we not doing about benchmarking that we need to address here? E.g., we say – what's a once in a 100 yr flux, and we put an error bar on it. CME isn't the worst-case scenario for everything. We have a once in a 100 yr storm, do we know what that means in terms of magnetospheric conditions? Not addressed sufficiently right now.

What do we use as our metrics for the benchmarks? Those that the users give us are not necessarily the metrics we want to use. Have we predicted the GIC at this particular location? It may be 50 km off or 20 mins off but the benchmark would not reveal that.

Uncategorized:

Focus on transformative, end users are important, certain things not well defined at present.

Modelers are not being paid, on a long-term basis, to develop models for forecasts. We can do the science of space weather, but not work on the transition to operations. And that aspect isn't changing. Maybe it is changing with the recent calls for R2O work. We have to show that this is a valuable process. If we want to make substantial progress, this work cannot be interrupted all the time. We now see NASA supporting this work with their Space Weather Science Applications program (SWxSA).

NOAA/NASA stuff was all O2R, which is not the same. R2O specifically not included in the DRIVE call. If we ask the NSF to pay for R2O, then who's going to pay for science?

Physics-based models are needed to understand how the predictions are being made. But in the end, these predictions may be made by machine learning or other data science techniques.

There is definite scientific value in running these models and having these standardized metrics that are being developed. That's something that our community should be doing, not just the modelers. We need to have the expertise, and standardization across models. This is not an operational activity, it's a scientific activity. Should not be considered operational.