2020 Verification workshop presentation notes (F.Gofa)

11 Nov 2020 – Session: HIW & USER VALUE CHAIN (Discussions with presenters)

Beth Ebert (BE): Overview of the HIWeather User-Oriented Evaluation Task Team Activities

Discussion

- How probabilistic is the value chain approach of warnings evaluation? Can people cope with that?
    - (BE): It depends on the crowd, the ultimate decision is deterministic of course. The trend is to become more probabilistic which we encourage it. Evaluating impacts is by its nature probabilistic.
- During COVID we have seen validation of the impact of community restrictions based on analysis of Google or Apple phone location data i.e. how much people are travelling. Has anyone looked at correlation between warnings and change of travel behavior as measured in this way?
    - (BE): There must be some research on that probably not in operational use. Travel patterns need to be looked in deeper, even if to have access to such data is not always easy.
- Curious if anyone is working on ontologies of these hiweather events? In our project we compared old and new snowstorms and found that we first had to define a storm that covered the full temporal extent of the storm (not necessarily easy to tell when the storm starts from sub-daily observational data). I'd argue that the need for ontologies is even more important when using social media data in unsupervised classification because you shouldn't rely on the unstructured content to determine what constitutes an event.
    - (BE): We will try to organize data in a way that will allow for ontologies to be done by collecting appropriate information. There is a new publication this year that classifies hazardous events and defines them, which can be a first step.
- The value chain seems to be the idea linking together all of these webinars -- they're at different locations along the chain. Would you agree?
    - Yes, absolutely. The value chain programs are presented in a linear flow but a lot of links go across different areas, different players, etc.
- There was an interesting study using this sort of data of how people responded to Superstorm Sandy in New York - distinguishing those who evacuated from those who stayed, and what each group were putting in their social media posts. (brianGold)
    - (BE): Interesting example that gives an opportunity to check what is working what not.

Barbara Brown (BB): User-driven evaluation of tropical cyclone predictions

Discussion

- Am I right in thinking that your metric for intensity is maximum wind speed? In the light of recently published research suggesting that total damage is poorly correlated with maximum wind speed, are you planning to look at other metrics of intensity, such as central pressure. which may be better correlated with damage?

    - (BB): Not doing this project anymore but that would be a good idea to include other factors apart intensity. There is lack of improvement in intensity forecasts that is why the focus was in this.
- The errors for the E1 and B2 were flat from about 48 hours on. Is that typical of intensity forecasts and if so why is that? Why don't we see an increase in error with lead time?
    - (BB): Need to look for the answer.
    - (Participant): Possible answer is that the maximum wind speed has a lot of noise in it and I suspect the errors are saturating by those lead times.
- Does that mean that you should use operationally different models on different lead time?
    - It is possible that if you had a really strong signal that you should. The goal was to show that overall there was not a degradation of forecasts. This is another quest to be asked by the value chain approach, the consistency of the forecasts.
- Tropical cyclone forecasting typically uses model consensus approaches. Did you do any evaluation of the extent to which a new model improved the quality of a consensus forecast (e.g. through having errors uncorrelated to the other members), as distinct from its individual performance?
    - (BB) There was some work done with the consensus forecasts. This would be a way to avoid switching back and forth between models, looking at the consensus. It always depends what is your focus in the analysis, what is your question.
- Operational forecasters can't look at an unlimited number of models. In the examples you showed it looks like they should add E1 and possibly stop using B1. Having E1 and B2, of similar standard seems more useful than just one of them. Any comments?
    - (BB): Goal was to select characteristics that overall suggest best forecast, best model. In general to Improve operational system.

Julia Chasco (JC): Incorporating the perspective of user evaluation into the creation of a new early warning system

Discussion

- Will the manual be available on the web?
    - There is a handbook for emergency agencies and for public available in Spanish.
- Do you have a fixed lead time for moving from Alerts to Warnings? I am thinking some phenomena can be Warned with a longer lead time than others. Will your maps show both Alerts and Warnings concurrently?
    - The maps provide information on both alerts and warning and the system will be available soon (…..)
- Traditional text based warning products are still useful, as was demonstrated in the New South Wales bush fires when some communities were cut off from outside communication apart from radio (radio presenters were reading out text-based warnings and advices prepared by BoM & emergency services). Has your team worked on ways to overcome difficulties in user interpretation of text based warnings?
    - One of the things identified in the text warning was the problem of technical language, especially when issued by emergency situation workers with no ability to communicate

differently. The new system allows for better outreach without cancelling what is existing.

- I like the social approach - do you have epidemiological input into the development of impacts as well?
    - No, we do not. For the heat wave early warning system was developed with heath ministry and has mortality impact information, in this early warning system epidemiological input is not included for now.
- The use of dashboard seems to be gaining popularity, can the users drill down information such as radar images complimentary to the EWS information?
    - Sure, this is already available. It is important if weather services accompany their information with such additional data.
- You mention the importance of building long term user relationships, how do you recommend maintain this effectively?
    - The pandemic context, did not make easy to work with emergency agencies and to train them in the EWS but the technology helped in this and will help more in the future, both to reach and train people but also to collect data. We have already a plan to reach users in 2021, in order to understand how they are using this EWS, we build a relationship by being there when they need.
- How did you overcome user survey fatigue given the multitude of surveys you ran? Did you offer chocolate?
    - Use different sources for questionnaires you do not address to the same crowd.

Amanda Anderson (AA): Verifying the Performance of a Coupled Fire-Atmosphere Model

Discussion

- WMO is thinking about whether exchange of hazard and impact data could be promoted with associated standards. Do you think this is possible?

It is possible and will be extremely useful. I'm not sure how long it will take or how difficult it would be. You would have to decide on a standard, how it would be implemented and who would follow it.

- Amanda, how is spotting handled in CO-FPS and in the verification? Is spotting treated any differently than the main fire spread region?

Spotting is handled through a module that was developed but it is not known if it is being used in the operational version. It is an operation that is based on the fuel, the fire behavious, and the wind were …. They have a module to handle that, particle tracing model and result in a likelihood map. Verification of a likelihood will be very difficult. Possible test case identified this year. Work still needed.

- What was the frequency of the stakeholder meetings? What was the duration? Who participated?

We had 2-3 stakeholder meetings per year (every 4-6 months) throughout the 5-year duration of the project, 4-5 hour duration and the participants are project managers from the Center of Excellence, fire-fighting community, regional and state level fire responders, air quality specialists and meteorologists predicting impact of smoke on air quality.

- How did stakeholder feedback / information advance your project?
    - It helped with the verification part and it was also integral in advancing the project as a whole and what areas of the science were focussed on. For example, the development of the spotting module was driven by the stakeholders. They also provided feedback about the forecast, if something was lacking or if something needed to be addressed.
- Does CO-FPS provide any information on the severity of the fire at a given point (e.g., ground fire, single-tree torching, intense crown fire)? If so, have you been able to do verification based on after-fire assessments of vegetation damage?
    - The main output is flame length, which itself was driven by stakeholder feedback Vegetation-based assessments of the model were not done, but video from aircraft was used to estimate flame length (e.g. from the height of a tree that was on fire). Fire cumulus were also examined as stakeholders pointed out that such cloud formation is a sign that the fire is about to become very unstable.
- Are video footage and photos from social media used as additional information to estimate the perimeter? I assume that information is often posted in real time.
    - Such media obtained from social media has not been used because it needs to have GPS data associated with it and very often this is not the case. However, social media has been used for studying air quality impacts, evacuations, and pyrocumulus cases.

D*iscussion and ending comments*

- (Participant) The US advisory/watch/warning system is one way of working. At the other end of the spectrum, UK warning levels are distinguished by risk, but all are called warnings, regardless of lead time.
- (JC) It really concerns each country what definition they give in warning or alert. There should be a protocol that follows any information. A weather service should work with a communication framework in order to bring to the public consistent and comprehensive information.
- (BE) Learn from EMMA, meteoalarm and such Europeans projects that they had to deal with this variability.