**Update on LPV Stages and Metrics for Cal-Val - Gabriela**

**(presentation link -** [**http://tinyurl.com/grlanv3**](http://tinyurl.com/grlanv3)**)**

The LPV validation stage table links to the validation paper we are writing.

Other similar approaches also linked to CEOS, and inconsistencies exist across these.

Do we want to merge, adopt another existing approach, …??

After updating the web site we wanted to write a paper that summarized validation stages for each variable. Determine where the bottlenecks are for these variables to reach the next stage.

In the process we discovered some inconsistencies with our current table.

Would like to publish in RSE, abstracts are due May 31.

GS posed the question to all, do you know the table, and if so, is the table important to you?

Consensus was yes. Most use or refer to it as a standard.

There was a comment from Luigi that for some categorical products the val stage definitions can have some unintended consequences. He stated that level 3 is not always better then level 2. For instance, a careful selection of a few representative sites at stage 2 is better than the same number of sites randomly selected that qualify as level 3 if the standard errors are large.

Something similar happens with SM. Small number of well-characterized sites vs. high density sampling doesn’t mean you are better than stage2.

Must see presentation (see link above) to follow the discussion about the other approaches and definitions.

Maturity matrix

GS: In polling the group she wanted to understand how quickly we can change things, how important is it to keep a certain level of consistency with the current table, or if it’s easy to switch to other approaches. And wanted to do so before introducing the other approaches to avoid bias.

**Bates Matrix**

There are several columns in the original Bates matrix.

Software Readiness, Metadata, Documentation, Product Validation, Public Access, and Utility.

Has 6 stages divided into 3 levels - 2 research levels and a services level, the last of which is where the user community can use the data, there are two rows at each level.

Paying attention to just the Validation column.

Stage 1 - little or none (which we do not have)

Stage 2 – minimal (doesn’t say much)

Stage 3 - uncertainty estimated for select locations and times.

Stage 4 – uncertainty estimated over widely distributed times/location by multiple investigators; differences are understood

Stage 5 – consistent uncertainties estimated over most environmental conditions by multiple investigators

Stage 6 – observation strategy designed to reveal systematic errors through independent cross checks, open inspection, and continuous interrogation; quantified errors

It goes up in a way that is very similar to what we have already. Ours are more descriptive and detailed than Bates.

**Core-Climax approach**

Based on Bates and Privette, but has been extended so it can be applied to other data records, including in situ and reanalysis data. We do have to consider this type of integration, especially if we want to use a similar table for our reference data.

They are extending the validation column to include uncertainty assessment more generally. LPV mostly covered physical consistency tests - looking for consistency in time series, performing intercomparisons, and direct validation with in situ data.

Core-Climax has same 6 columns, except one, which is called uncertainty characterization. What is nice about it is how it builds up, it’s very consistent.

The uncertainty characterization column is actually filled with several sub-criteria (standards, validation, uncertainty quantification, and automated quality monitoring).

May need to include open access to the reference data, not currently addressed.

It would be good to have a new approach that fits into the maturity matrix. Do we use Bates or Core-Climax. Split it into subcategories, make a summary and put that into the maturity matrix.

Inconsistencies with our current table.

1. missing an indication that stages build upon each other.

->include

1. sites

a bit of a mess, stage 1 says you can’t go beyond level 1 with less than 30, then stage 2 just says a ‘significant’ number of sites, which is meaningless, and then in stage 3 it says multiple locations representing global conditions. Should this be variable-specific?

MR: sees stage 3 as implying representative biomes and seasonal conditions.

GS: we need to be general enough to cover all variables, we have to refer to the specific protocols, to be variable-specific for the stages.

MR: this would mean 9 different tables.

GS: Can we come up with a general statement that is good enough to cover all variables?

MR: that is what WGCV is asking us to do.

GS: do we base this criteria on statistics, or make our approach based more on ecosystem distribution ?

MR: digging up Privette’s old NPOESS approach that categorized validation into 4 different approaches

* point-based (BELMANIP – distributed sites) – land albedo, LST, VI, SR, snow
* remote sensing - things like ice and albedo in the cryosphere – in situ augmented by aircraft data
* episodic validation – done for targets of opportunity (fires)
* land cover typed RS – leverages hi-res data tasked for a site, data are independently classified and validated first

MR: One table per ECV would be too prescriptive, but we can use more broad categorizations for multiple ECVs. These categorizations would be helpful for earth science in general, because episodic isn’t just fire, it’s also applications (drought, floods, LCC). This is a way for us to go beyond variables, into adaptation. It would show that we are focusing on earth science data needs for accuracy. A few broad categories that would encompass multiple ECVs.

There is an advantage to not becoming too specific, such as prescribing a number of sites.

We should stick to 3 categories, BELMANIP approach, RS approach, espisodic approach.

LB: Suggests continuous variables, categorical variables, categorical change (ephemeral change of a categorical variable or a permanent change of a categorical variable), all need slightly different sampling approaches and different ways of assessing the accuracy.

GS: It seems clear that we have to make a new version of the table, from the discussion and statements it is understood that we can’t just totally change it. This will have implications for the agencies, and the science teams, so we have to be careful.

**Part 2: (after lunch)**

Looking back at the tables.

Discussion of the levels, and the fact that Bates has a non-validated level.

GS: These levels are applicable to CDRs, there should be at least some validation in order for something to be called a CDR, no?. Discuss w Yurk (?).

What about provisional products, or experimental products, such as the L8/S2 products. These are not validated at all. Suggestion that rather than ignoring or not talking about it, these types of products should be given a level. And the suggestion is that the levels should not start at 1! This is commonly used as a base stage, with some validation. Zero designation would be better.

GS: Yurk(?) – also suggested that the level 4 that we have, where it applies to when new versions come out or as time series is expanded. When a new version comes out, this would be a new maturity matrix. A new stage assignment for the new product version. As we automate validation, periodic checking and update to make sure that as the series expands these data are checked against new reference data that become available.

GS: Summary – Bates and LPV really just addresses validation, Core-Climate goes beyond, including uncertainty assessment. Bates and LPV both missing any uncertainty characterization of the reference data.

Core-Climate based on sub-criteria columns - GS recommends this is the way to go for LPV as well.

We know we have inconsistencies, and we have to update the validation stage table.

Question at WGCV level. Should come up with a generic table they can use as well that would cover things at the WGCV level, such as atmosphere.

Need also to make it consistent with other existing approaches, such as the Maturity Matrix. Best to make them compatible going forward, so we don’t have to revisit.

MR: Re WGCV, preference is for us to finalize our tables first. We will refine them in such a way that meets the needs of different subgroups.

- need to add something w/ respect to stage 0

- need to add implied requirements that are not there

An update will help address some of the concerns from the other groups, there will be push back, resistance, but GS states, that if we add sensible subcategories they can replace terms as needed and that will be helpful for all.

GS: Metrics – not really addressed in presentation, uncertainty terminology is separate from our validation stage table. This was a task of IVOS (Nigel) but it didn’t happen. We shoudn’t take it on. GUM not always applicable to the validation of land products. This is another problem.

Should be an action, that somebody presents a certain set of matrices.

Discussion?

Are we missing any tables?

Should we adopt a Maturity Matrix? Bates? Core-Climax?

Validation – for us there are 3 components

1 – physical consistency – plot time series, look for consistency within a product (as in OLIVE) – not really validation – this assures everyone does it

2 – intercomparison between different products

3 – direct validation with in situ data

MR: says that 1 is QA – consistency check is something you need to do throughout the entire mission, it’s part of maintaining a standard product. This is something that MODIS/VIIRS teams do as part of their procedure. Not everyone does this, it may be implied, true its QA not validation, but we need to leave it so it’s understood.

Seems we are heading toward a direct consolidation with the maturity matrix.

The Paper –

Not sure we are ready.

Important to have authors from different agencies and those with different applications in mind as we go into developing the table.

Need to work with WGCV and the climate community.

We come up with a good idea, distribute and go from there.

Who leads, contributes?

Need to agree on the categories, how many validation approaches are we going to slice this in to, and then for task groups to agree on the stages.

We want this to scale into other areas of RS (one for L1, one for SAR, each w/ their own subcategories).

Perhaps we to follow up - get a group of 5-6 people who go though the documents, and then have a telecon and then define the categories.

Can we define the core categories for the approaches now?

How many tables are we doing?

Type of variable (continuous, discrete, etc)

This then evolved into a long discussion, of the tables and categories ….

How do in situ data evolve through the stages, this should also be stated.

Traceability, characterizing the reference data

List of columns to characterize the properties of the validation process.

How many stages?

We will have stages 0-4.

Their level of maturity 1-2 is our level 0

* is the product accurate? At this stage it is still unknown, still doing QA, stabilizing the instrument, not in evaluation mode yet.

We discussed many times having multiple tables for different types of variables (discrete, continuous, …etc).

MR suggested that we do not categorize, but rather expand the columns to accommodate these.

SP suggested we do not change the (MM) table, the table is there in place, try to work with it. Everyone agreed that part of the strength of our table was its simplicity. GS also offered that Nigel suggested we just give additional explanations where needed to go with the table. Address different validation types, keep as generic as possible, include case studies.

MR suggests we create an updated version of the stages on the web site that retains the validation stages that we have, adds stage 0, attach a separate section that list the types of validation that are acceptable, and show how they match up to the stages.

Still need to address the sage 3 to stage 4 inconsistencies.

Suggestion that validation and uncertainty characterization difference is not completely clear in the core-climax table.

Need to push fiducial reference data into the standards, SI traceability. Open access. Miguel suggests that open access does not need to be addressed specifically, he says that by using the fiducial term that this is implied in the definition. By using this term it is implied that it is both SI traceable and open.

The validation column of Core-Climax covers the temporal and spatial aspects, as does ours.

Actions:

1 – revise LPV val stages to reflect stage 0

2 – revise our set of level 3 and 4 definitions to better account for the need for fiducial reference data and how that is taken into account.

3 – add a section of validation approaches and case studies to inform the user as to proper approach to validation for each type of validation (specific examples for fire, BA, .. all categorical variables. i.e. BA cannot have a fiducial measurement.

(Q: This effort may partly address Chris’ comments about validation principles)