Science Team Proposal for Philosophy and Structure of Discharge Data Products

Colin Gleason, Michael Durand,
Pierre Olivier Malaterre

SWOT Science Team Meeting,
Bordeaux. June 18, 2019

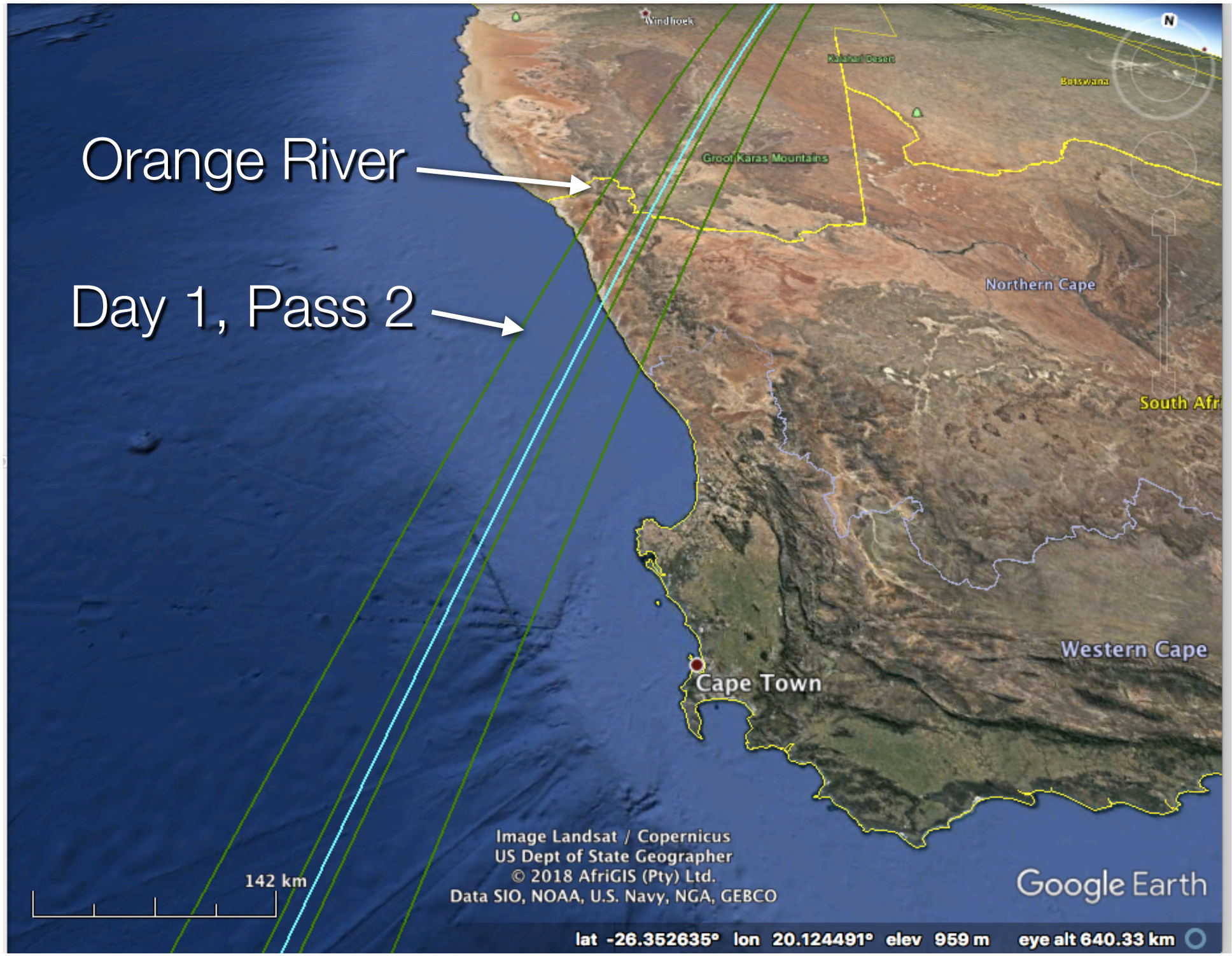
Photo by Rob Krummeck, via Google Earth
The R358/C10 bridge over the Orange River

1. Background: data product structure. DAWG & Project.
2. Discuss philosophy of data product: a priori data

Colin Gleason, Michael Durand
Pierre Olivier Malaterre

SWOT Science Team Meeting,
Bordeaux. June 18, 2019

Photo by Rob Krummeck, via Google Earth
The R358/C10 bridge over the Orange River



Motivation: ~April 2022, the first SWOT science orbit will measure a river

The discharge algorithm working group will work with the project to create the discharge data product

Paradigm: Science team estimates parameters. Project computes discharge

- "Official" discharge product to use simple flow laws and uncertainty models e.g.:

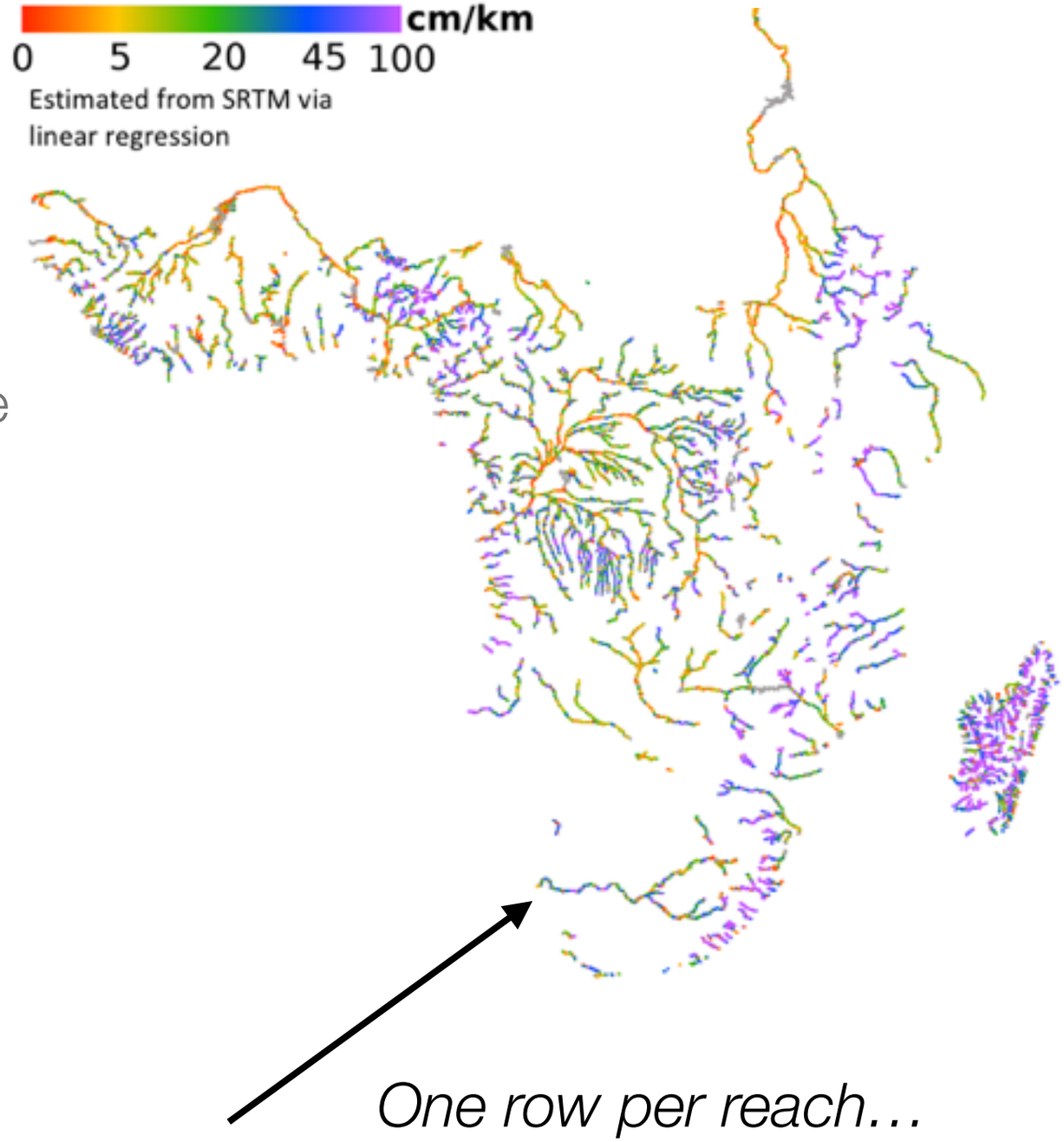
$$Q = \frac{1}{n}\left(\bar{A} + A'\right)^{5/3} W^{-2/3} S^{1/2}$$

- where n is roughness coefficient, and $\bar{A}$ is time series median cross-sectional area (over DAWG-specified inversion window): **not observables**

- Science Team computes parameters, using McFLI, VDA, integrators, etc.

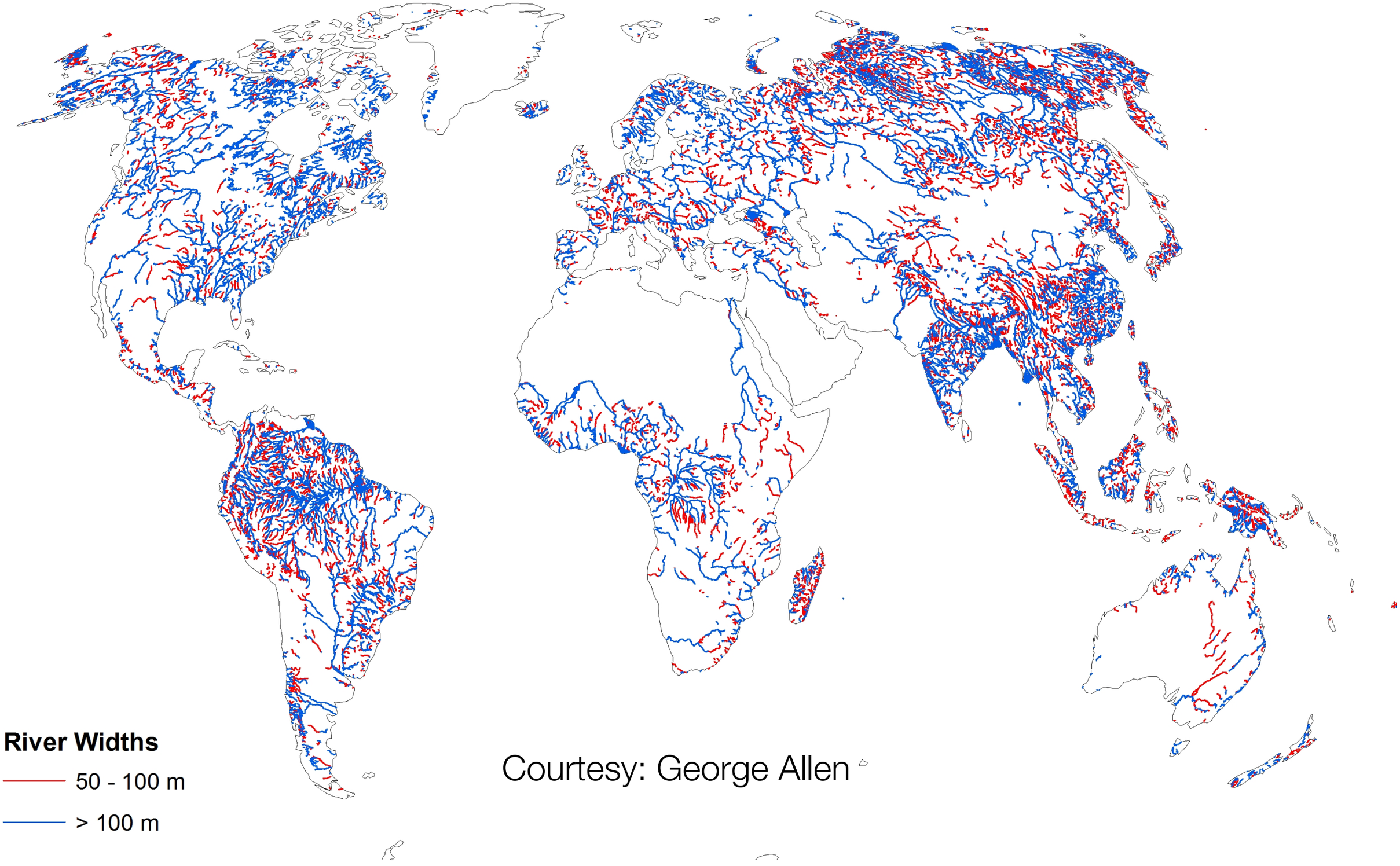
- Multiple algorithms will be included, along with a a "consensus" TBD algorithm

- All of these must go through QA/QC



One row per reach…

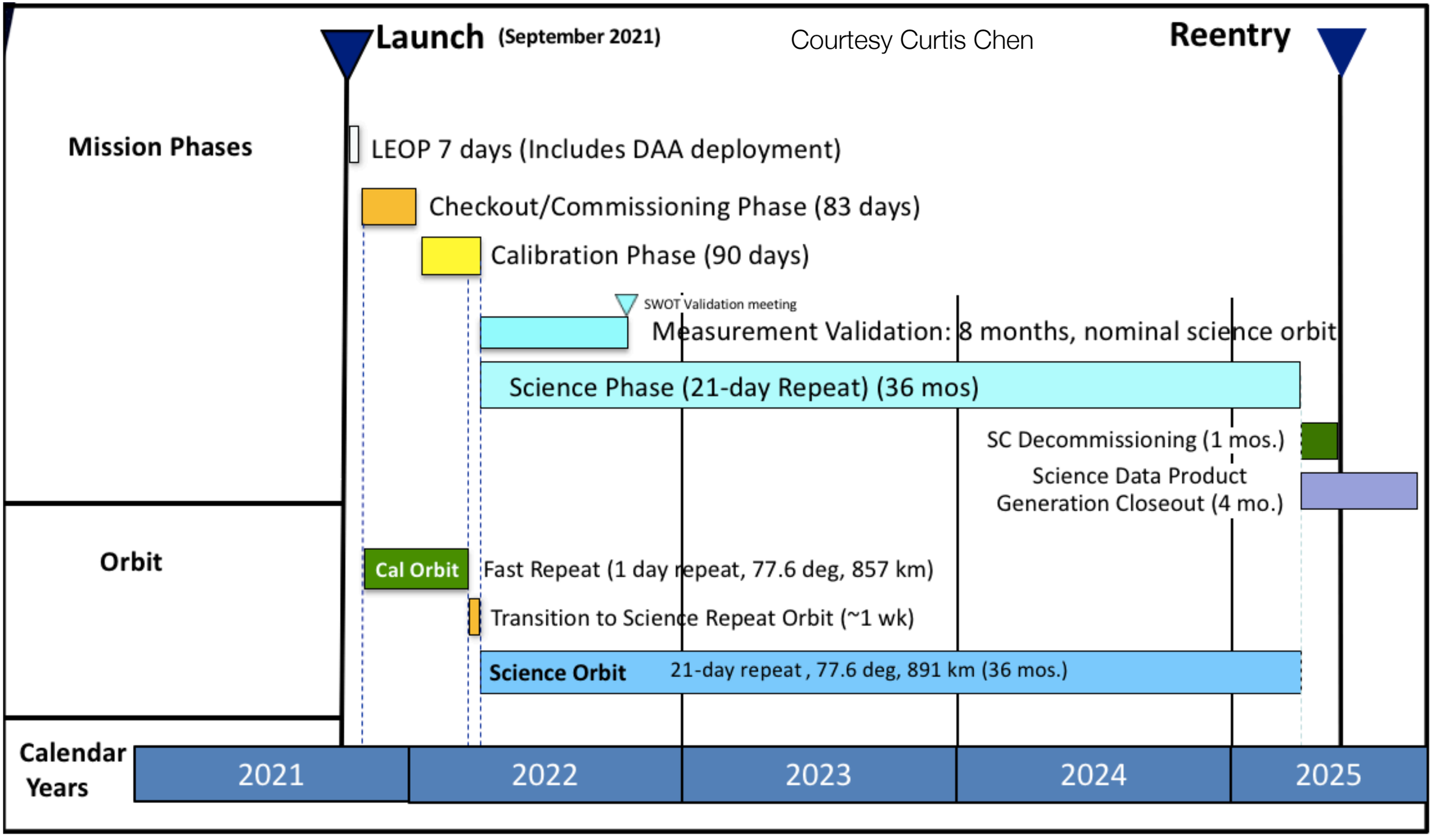
Id	Alg1 n	Alg1 $\bar{A}$	Alg2 n	Alg2 $\bar{A}$	…
1					
2					
…					
n					



Courtesy: George Allen

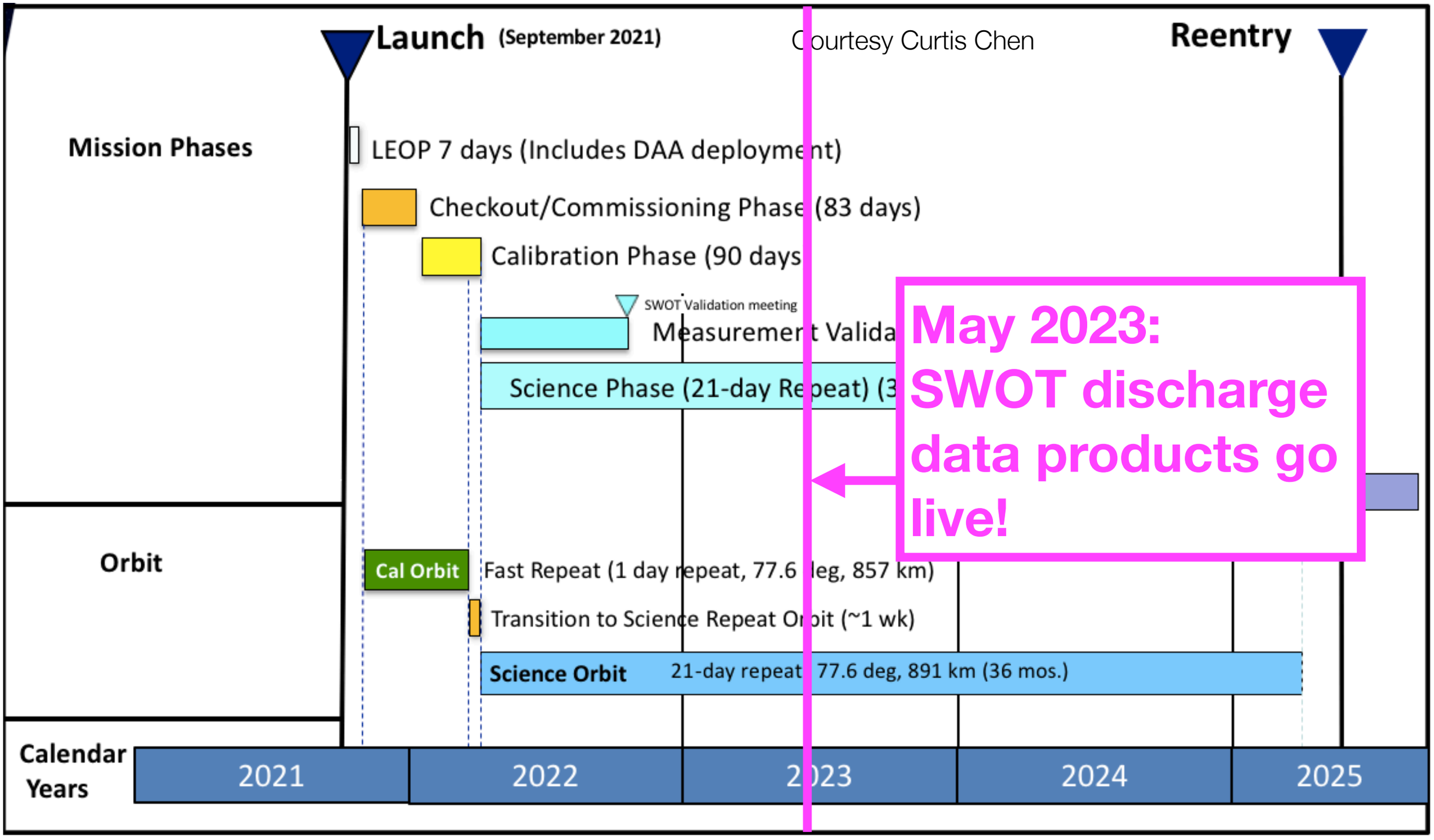
Parameters must be estimated for all reaches

Optimally computing these parameters is a heavy lift but doable. See talk on “Confluence” project by C. Gleason



Mission Timeline

The DAWG will use science orbit data to compute optimal parameters



Discharge data product timing specified in SRD

Data products to be produced ~ one year into science orbit

Today’s goal is to present and discuss a proposal for discharge product use of a priori data and (thus) product philosophy

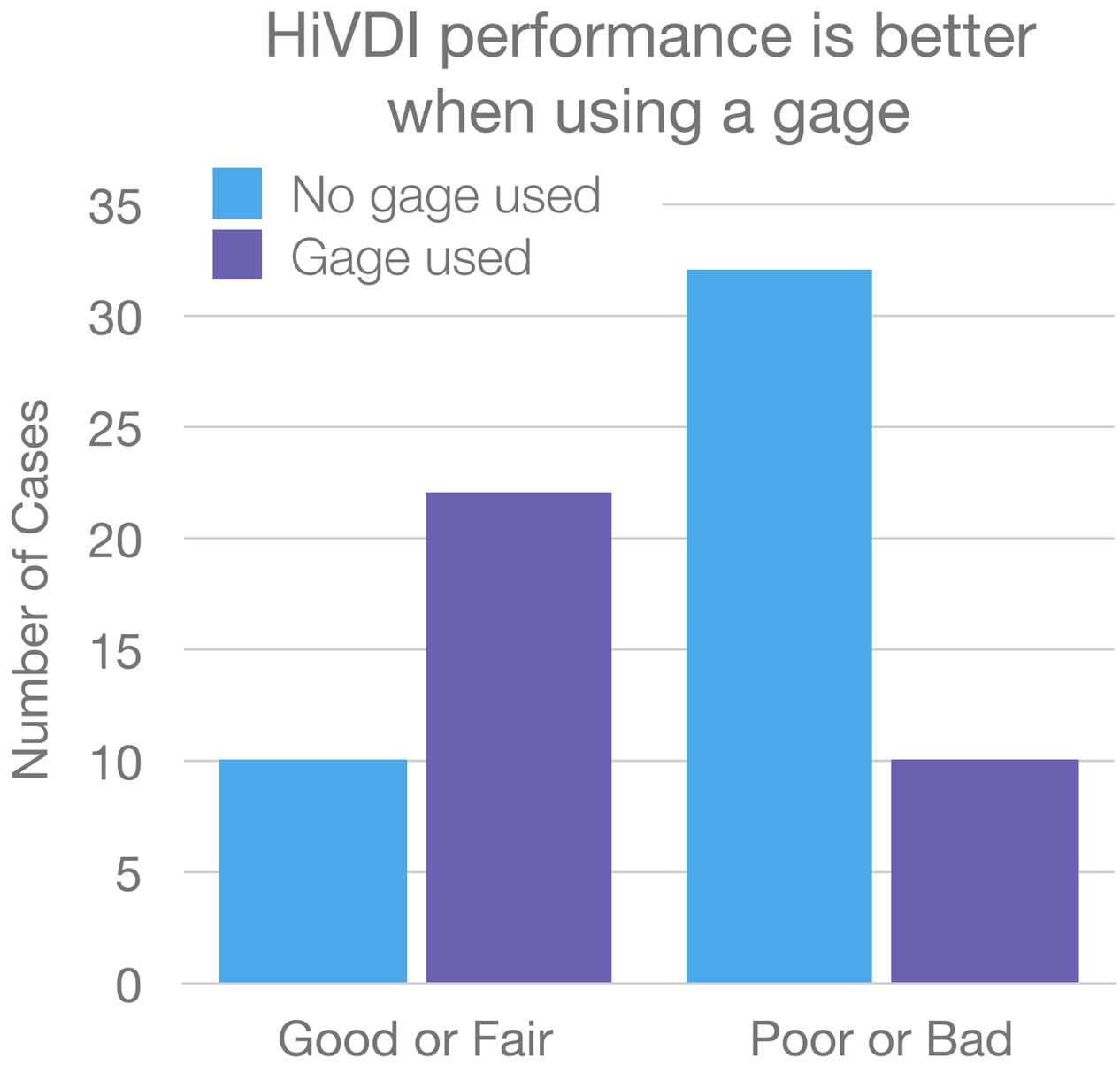
Photo by Rob Krummeck, via Google Earth
The R358/C10 bridge over the Orange River

What is the best way to use prior data, e.g. real-time gages in SWOT reaches?

- Pepsi2 has shown that discharge accuracy quality partly controlled by a priori data (e.g. see HiVDI results)

- What to do in cases where SWOT data overlaps ~real-time discharge gages?

- Need philosophy for using a priori data.



Proposed philosophy on the use of a priori data

- The final data product should be as accurate as possible, and must have validated uncertainty

- Thus, the consensus SWOT discharge data product **should use all types of a priori data**, including real-time data.

- We must separate out a subset of in situ data (**not to be touched** in creating data product) to assess product accuracy, and give confidence to product uncertainty estimates

- The **DAWG must develop a database of discharge records**, linked to the SWOT a priori river database (SWORD).

Proposed philosophy on the use of a priori data

- All algorithms must be absolutely clear on what a priori data is used.

- Algorithm uncertainty estimates should be informed by what a priori data is being used

- All algorithms must have access to all discharge records

- Something to consider: produce a “no real-time gage data” product, included as an “expert” product? Note Pierre Olivier Malaterre is in favor of this

Discussion points

- We assume we agree on the background: the discharge algorithm paradigm and basic plan as laid out at the 2018 ST meeting Discharge Workshop, and the implications for the next ST effort required

- Let’s discuss philosophy of how to use a priori data, using the proposal on the previous slide as a jumping off point.

- If there’s time, we can discuss the laundry list of topics for the next ST to sort out.

Laundry list of TBD topics

List to organize, not to discuss today

- How will QA/QC of discharge parameters be performed?
- Consensus algorithm development and testing
- Cross-sectional area reference: median vs low flow
- What other prior data needs to be assembled?
- We need to assess discharge uncertainties

Propose that next ST hold a DAWG workshop ASAP after start of the next ST meeting to make decisions on these issues

Backup Slides

Photo by Rob Krummeck, via Google Earth
The R358/C10 bridge over the Orange River

The Science Team vis-a-vis the Project

- The implications of the proposal on previous slide are important. It will take significant effort to produce and validate these sets of parameters

- We propose that the Science Team keep space for this work in the next round of funding. The “science” to be done is twofold:

    1. Developing and implementing methods for producing discharge data product parameters, and

    2. Evaluating how the data product contributes to understanding of global water cycle science

There are three classes of available methods

- McFLI and Variational Data Assimilation (VDA) have (thus far) been demonstrated at reach scale, and have significant history (~20 publications?)

- So-called “Integrator” methods are newer, and involve methods to i) merge non-SWOT observations with SWOT discharge; and ii) ensure consistency of discharge estimates across river confluences and networks. Much less mature.

- There are endless permutations and options for combining McFLI, VDA, and Integrators. **How to best deploy these methods for SWOT discharge?**

The endpoint we are considering today is the discharge product computed by the SWOT Project

- The SWOT Project discharge data product will be available with the river data products. It will be evaluated in some sense by the project. It will be global.

- The Project discharge will consist of: 1) Individual discharge algorithm estimates (five currently proposed) computed by individual algorithms such as BAM+U Mass Integrator. 2) A “consensus” algorithm estimate that gives an estimate based on the individual algorithms.

- Note the difference between Project discharge data product, produced by the project, and science team data products where you have flexibility to put anything you want on a website. Here we’re focused only on the Project product.

Proposed algorithms to include in official data product

1. BAM (to be expanded to include a routing integrator): U Mass - Gleason

2. HIVDI: Strasbourg

3. MFGA: USGS

4. MetroMan (to be expanded to include a linear integrator): OSU

5. SADS: U Mass (Andreadis)

More proposals expected, e.g. Montpelier

Some specifics: Algorithm components in the SWOT a priori river database

- Each of these five algorithms will have a basic set of parameters, e.g. roughness coefficient and river cross-sectional area at median flow stored in the “SWOT a priori river database”. Each of the ~200,000 reaches will have its own set of values for each each algorithm.

- Algorithms may use whatever prior data they want, but need to be 100% transparent about what is used, and how it is used

- There will be a consensus algorithm that will choose a set of parameters that will aim to predict a median discharge across all algorithms. These parameters will also be in the database. This presupposes choosing a form of Manning’s equation or some other flow law for the consensus algorithm.

How to balance goals of SWOT discharge product with the availability of in situ data?

- Do we incorporate non-SWOT information in the SWOT discharge product? To what extent? How do we handle this from a philosophical point of view?

- Easy example: at a confluence, easy to have imbalance if we just use McFLI. Provide constrained discharge? Or use integrator to smooth out these issues and provide mass-conserved discharge at network scale?

- Harder: when a reach contains a real-time discharge gage, do we really want McFLI-type accuracies, i.e. ±40% error? Or should we incorporate in situ data into our estimates?

A few important specifications: many more will be added in future

- Not all algorithms have to run everywhere
- All discharge algorithms need to have uncertainty estimates
- All algorithm parameters will be sent to the ADT for validation (probably simple boundary checks etc) before being sent to JPL, and finally merged with the river database and used online.