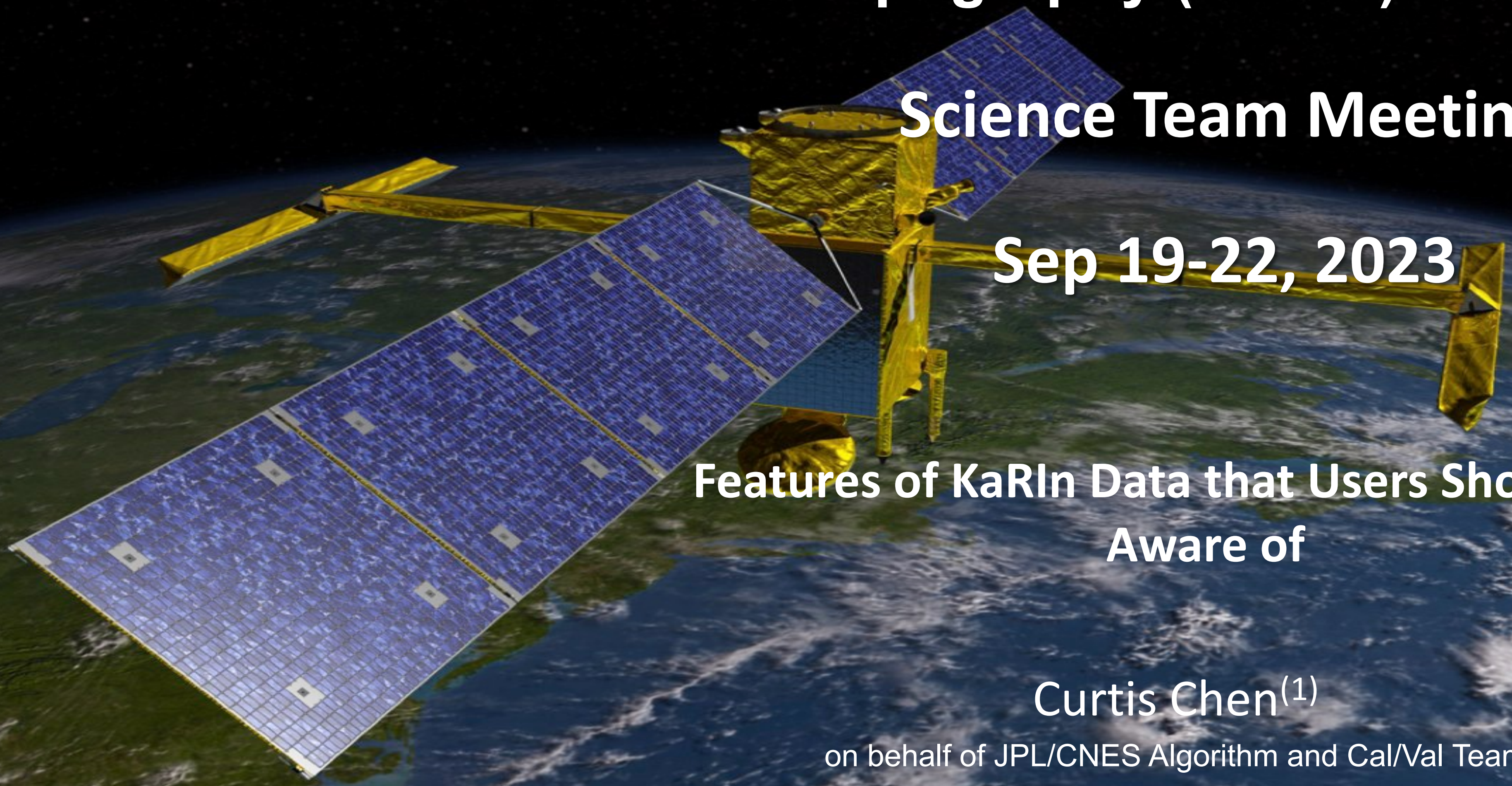
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Surface Water and Ocean Topography (SWOT) Mission



Science Team Meeting

Sep 19-22, 2023

Features of KaRIn Data that Users Should be Aware of

Curtis Chen[1]

on behalf of JPL/CNES Algorithm and Cal/Val Team

[1]Jet Propulsion Laboratory, California Institute of Technology

Introduction

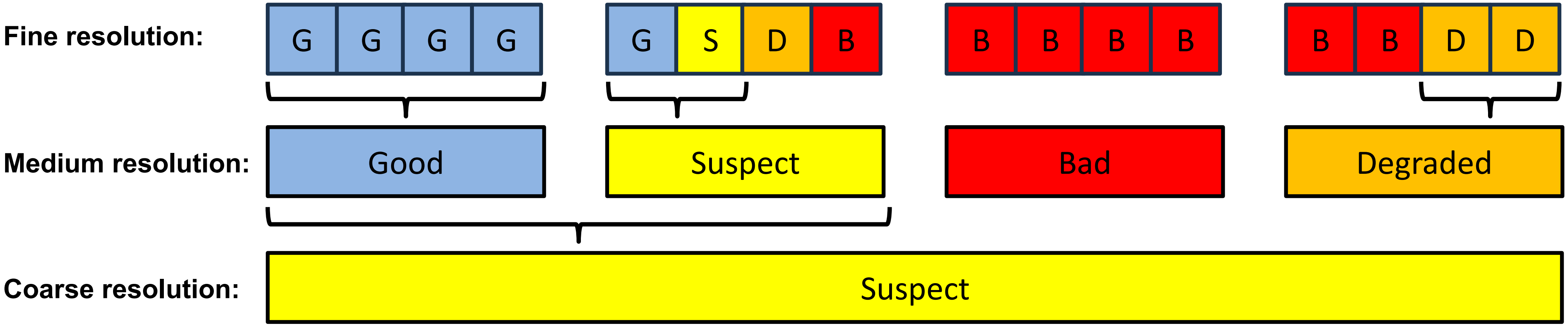
- KaRIn measurement is complicated!
  - Data products attempt to abstract complexities of measurement from users as much as possible, but many items that may not be immediately intuitive remain
  - Knowledge of measurement details can be especially important in trying to interpret pre-validated data products
- This talk addresses practical aspects of interpreting KaRIn data products
  - Answers to frequently asked questions
  - Tips to hopefully avoid misinterpretation and confusion
- General topics:
  - Definitions, conventions, and data representation
  - Data availability
  - Phenomenology to be aware of

Look at the Quality Flags!

- **Users should pay attention to quality flags in KaRIn products**
  - Measurement values are associated with quality flags
  - Quality flag variables are usually called `{measurement_variable}_qual`
    - Example: If measurement variable is named `height` then associated quality flag is usually named `height_qual`
    - Quality flag variable name for given measurement variable is indicated by metadata in product
      - NetCDF: See the *quality_flag* variable attribute
      - Shapefile: See the *quality_flag* field in the shp.xml file
- Quality flag indicates whether measurement is "good," "suspect," "degraded", or "bad":
  - "Good": Processing did not find any reason to disbelieve measurement
  - "Suspect": Something about measurement was not quite as expected, so measurement may be worse than normal, but may also be fine
  - "Degraded": Something about the measurement was definitely wrong, so measurement is likely worse than normal (though not necessarily by a lot)
  - "Bad": Measurement is likely nonsensical (e.g., null filled)

Quality Flags and Averaging

- Quality flags affect how averaging is done during ground processing
  - Good and suspect data samples are always used when averaging
    - If only good samples are used, then averaged output is marked good
    - If suspect samples are used, then averaged output *may* be marked suspect
  - Degraded data samples are used when averaging only if there are too few good and suspect samples
    - If degraded data samples are used, then averaged output is flagged as degraded
  - Bad data samples are never used when averaging



Quality Bit Flag Interpretation

- Many quality flags in different KaRIn products are defined as bit flags
  - Bit flag is unsigned integer whose individual bits indicate different off-nominal conditions
  - Allows single variable to contain multiple levels of information
- Simplest interpretation: See if flag is 0 or nonzero
  - If flag = 0, then measurement is "good" (0 always means "good" for KaRIn quality flags)
  - If flag $\neq$ 0, then measurement is not "good"
- Straightforward interpretation: Interpret flag as numeric value and compare to Threshold1 and Threshold2 defined in metadata of flag variable (and PDD):
  - If flag = 0, then measurement is "good" (0 always means "good" for KaRIn quality flags)
  - If flag $\neq$ 0, then
    - If $0 <$ flag $\leq$ Threshold1, then measurement is "suspect"
    - If threshold1 $<$ flag $\leq$ Threshold2, then measurement is "degraded"
    - If flag $>$ threshold2, then measurement is "bad"
- More sophisticated interpretation:
  - If flag = 0, then measurement is "good" (0 always means "good" for KaRIn quality flags)
  - If flag $\neq$ 0, then decompose nonzero flag value into individual nonzero bits to determine what exactly was not good about measurement (see example on next slide) and interpret measurement accordingly
    - Top-level bit definitions are in metadata of flag variable
    - Additional details on bit definitions are in PDDs

Quality Bit Flag Example

See Appendix B of L2_LR_SSH PDD for details on this example

Bit (from LSB)	Decimal	Hexadecimal	ssh_karin_qual	ssha_karin_qual	swh_karin_qual	sig0_karin_qual	wind_speed_karin_qual
0	1	1	suspect_large_ssh_delta	suspect_large_ssh_delta		suspect_large_nrcs_delta	
1	2	2	suspect_large_ssh_std	suspect_large_ssh_std		suspect_large_nrcs_std	
2	4	4	suspect_large_ssh_window_std	suspect_large_ssh_window_std		suspect_large_nrcs_window_std	
3	8	8	suspect_beam_used	suspect_beam_used	suspect_beam_used	suspect_beam_used	suspect_beam_used
4	16	10	suspect_less_than_nine_beams	suspect_less_than_nine_beams	suspect_less_than_nine_beams	suspect_less_than_nine_beams	suspect_less_than_nine_beams
5	32	20			suspect_rain_likely		
6	64	40	suspect_ssb_out_of_range	suspect_ssb_out_of_range			
7	128	80	suspect_pixel_used	suspect_pixel_used	suspect_pixel_used	suspect_pixel_used	suspect_pixel_used
8	256	100	suspect_num_pt_avg	suspect_num_pt_avg	suspect_num_pt_avg	suspect_num_pt_avg	suspect_num_pt_avg
9	512	200	suspect_karin_telem	suspect_karin_telem	suspect_karin_telem	suspect_karin_telem	suspect_karin_telem
10	1024	400	suspect_orbit_control	suspect_orbit_control	suspect_orbit_control	suspect_orbit_control	suspect_orbit_control
11	2048	800	suspect_sc_event_flag	suspect_sc_event_flag	suspect_sc_event_flag	suspect_sc_event_flag	suspect_sc_event_flag
12	4096	1000	suspect_tvp_qual	suspect_tvp_qual	suspect_tvp_qual	suspect_tvp_qual	suspect_tvp_qual
13	8192	2000	suspect_volumetric_corr	suspect_volumetric_corr	suspect_volumetric_corr	suspect_volumetric_corr	suspect_volumetric_corr
14	16384	4000					
15	32768	8000	degraded_ssb_not_computable	degraded_ssb_not_computable			
16	65536	10000	degraded_media_delays_missing	degraded_media_delays_missing		degraded_media_attenuation_missing	degraded_media_attenuation_missing
17	131072	20000	degraded_beam_used	degraded_beam_used	degraded_beam_used	degraded_beam_used	degraded_beam_used
18	262144	40000	degraded_large_attitude	degraded_large_attitude	degraded_large_attitude	degraded_large_attitude	degraded_large_attitude
19	524288	80000	degraded_karin_ifft_overflow	degraded_karin_ifft_overflow	degraded_karin_ifft_overflow	degraded_karin_ifft_overflow	degraded_karin_ifft_overflow
20	1048576	100000					
21	2097152	200000					
22	4194304	400000					
23	8388608	800000					
24	16777216	1000000	bad_karin_telem	bad_karin_telem	bad_karin_telem	bad_karin_telem	bad_karin_telem
25	33554432	2000000	bad_very_large_attitude	bad_very_large_attitude	bad_very_large_attitude	bad_very_large_attitude	bad_very_large_attitude
26	67108864	4000000		bad_tide_corrections_missing			
27	134217728	8000000	bad_ssb_missing	bad_ssb_missing			
28	268435456	10000000	bad_radiometer_corr_missing	bad_radiometer_corr_missing		bad_radiometer_media_attenuation_missing	bad_radiometer_media_attenuation_missing
29	536870912	20000000	bad_outside_of_range	bad_outside_of_range	bad_outside_of_range	bad_outside_of_range	bad_outside_of_range
30	1073741824	40000000	degraded	degraded	degraded	degraded	degraded
31	2147483648	80000000	bad_not_usable	bad_not_usable	bad_not_usable	bad_not_usable	bad_not_usable

Final Note About Quality Flags

- Flagging algorithms are complicated and still evolving
  - New flag bits may be defined
  - Internal thresholds for when to set different flag bits will likely be adjusted
  - Possible that flags may be removed or repurposed
    - Especially if those bits are never raised
  - Threshold1 and Threshold2 for determining suspect vs. degraded vs. bad may change in future product versions
    - But flag interpretation will always remain consistent for a given version
- Flags themselves may have bugs
  - Some known bugs in flags are in “beta pre-validated” products
    - That’s why the products are “beta pre-validated”
  - Project still recommends that users look at flags first

Uncertainty Estimates

- Many measurement variables are accompanied by uncertainty estimates
  - Uncertainty variables are usually called `{measurement_variable}_uncert`
    - Example: If measurement variable is named `height` then associated uncertainty estimate is usually named `height_uncert`
  - Uncertainty estimates are typically 1-sigma (68th percentile) values
  - Uncertainty estimates often reflect only random error, not systematic error
  - See PDDs and ATBDs for details
- Uncertainty estimates for KaRIn height estimates are usually based on interferometric coherence
- **Validation of uncertainty estimates has been lower priority than validation of measurement variables themselves**
  - Use with caution
  - Do not be surprised if observed systematic errors exceed uncertainty estimates significantly

L2_LR_SSH "_2" Variables

- L2_LR_SSH product has two versions of SSH and SSHA with different wet troposphere and sea state bias (SSB) corrections:
  - Solution 1 relies more on observations for corrections (may be slightly more accurate)
  - Solution 2 relies more on models for corrections (fewer gaps due to missing corrections)

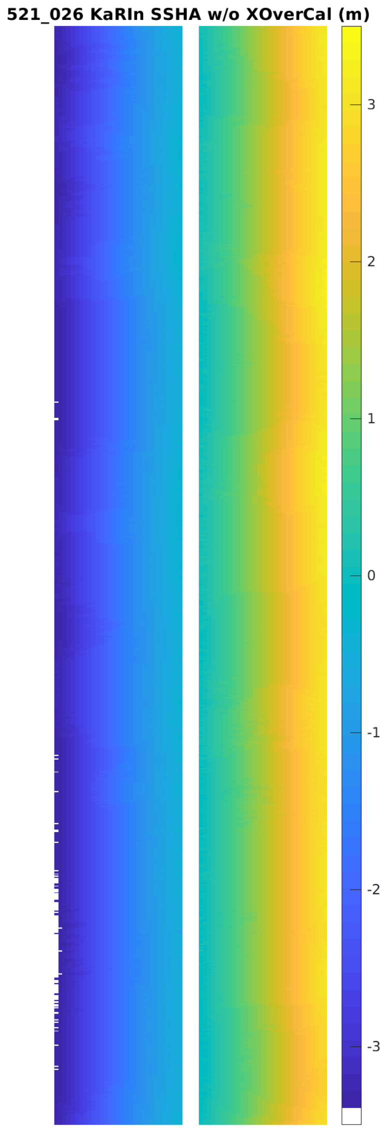
Property	Solution 1	Solution 2
SSH variable name	`ssh_karin`	`ssh_karin_2`
SSHA variable name	`ssha_karin`	`ssha_karin_2`
Wet tropo correction source	Radiometer	ECMWF model
Variable indicating wind speed used for SSB correction	`wind_speed_ssb_cor_source`	`wind_speed_ssb_cor_source_2`
Wind speed source as of Sept 2023	KaRIn for beta-pre-validated release Nadir altimeter for pre-validated release	ECMWF model
Variable indicating SWH used for SSB correction	`swh_ssb_cor_source`	`swh_ssb_cor_source_2`
SWH source as of Sept 2023	Nadir altimeter*	Nadir altimeter* for beta-pre-validated release ECMWF model for pre-validated release

**Nadir altimeter SWH is smoothed before SSB computation*

- **Input info for SSB correction solutions may continue to change in future product releases**

Application of Crossover Calibration Correction

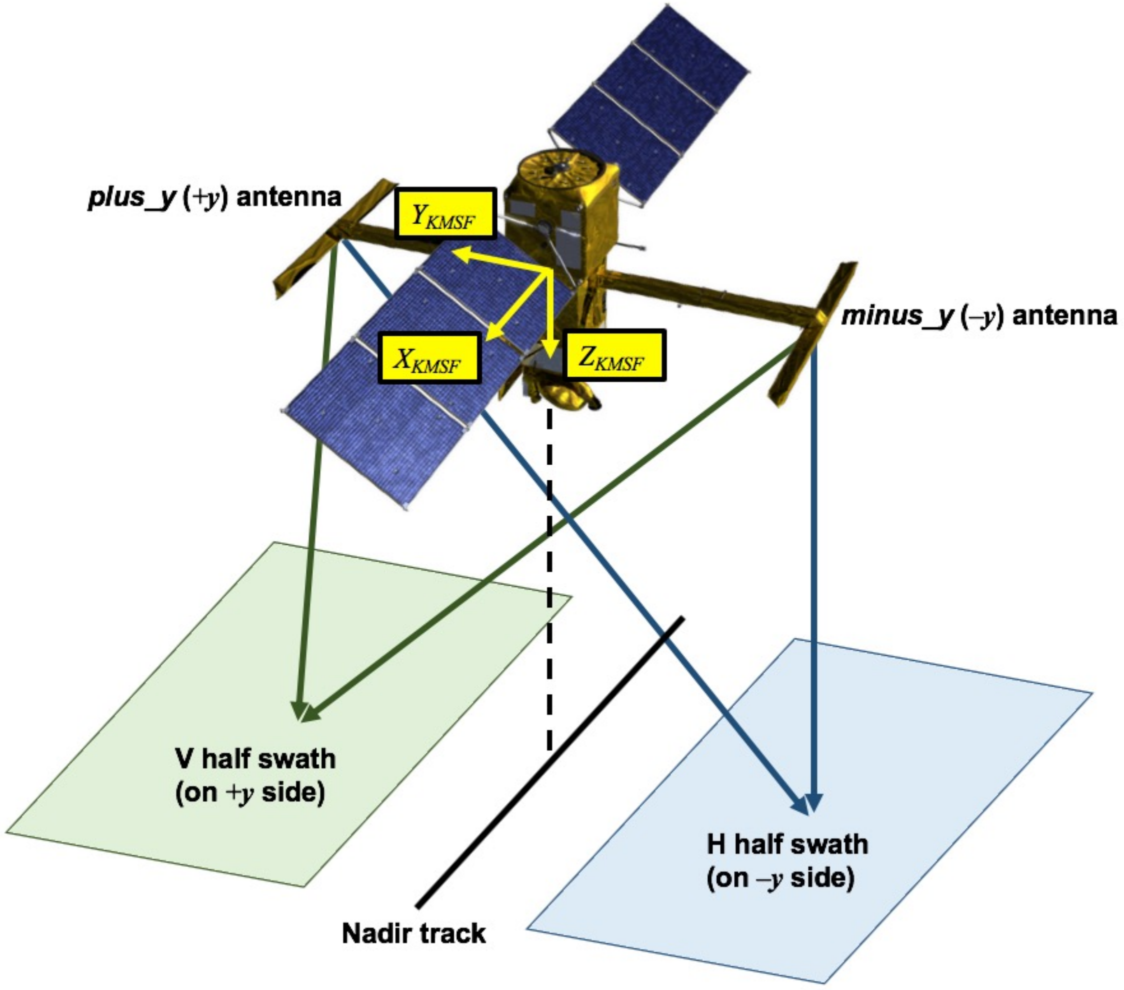
- Correction from crossover calibration (also called XOverCal or “xover”) is reported in L2_LR_SSH product but ***is not*** applied to SSH or SSHA
  - To get corrected SSH, user must compute following himself/herself:
    - `ssh_karin_corrected = ssh_karin + height_cor_xover`
    - `ssha_karin_corrected = ssha_karin + height_cor_xover`
    - `ssh_karin_2_corrected = ssh_karin_2 + height_cor_xover`
    - `ssha_karin_2_corrected = ssha_karin_2 + height_cor_xover`
  - Crossover correction has its own quality flag in L2_LR_SSH
    - Example: If considering `ssha_karin_2_corrected` above, should examine both `ssha_karin_2_qual` and `height_cor_xover_qual`
- Crossover correction ***is*** applied to height in L2_HR_PIXC and thus also to water surface elevation (WSE) in RiverSP/Avg, LakeSP/Avg, and Raster products.
  - Crossover quality is indicated by bits in relevant height or WSE quality flags in HR products
  - If crossover correction quality flag indicates bad correction, then PIXC result is flagged as “degraded” and uncorrected height is reported



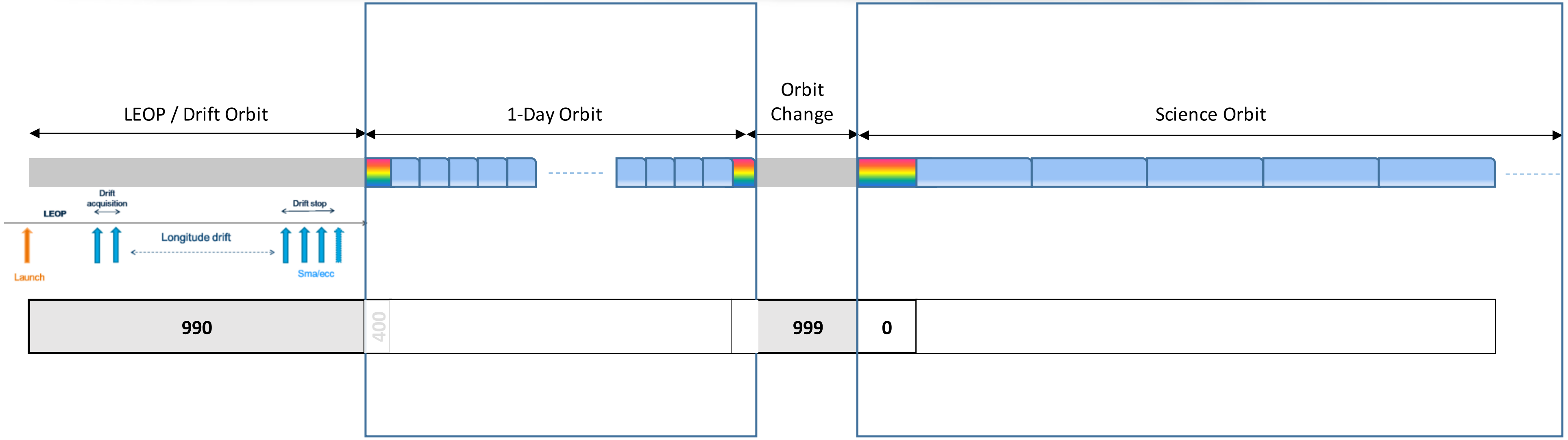
KaRIn SSHA directly from L2_LR_SSH product shows +/-3 m tilt in cross track without XOverCal correction

Left, Right, H, V, Plus Y, and Minus Y

- SWOT spacecraft undergoes 180˚ yaw flips every ~2.5 months
  - Spacecraft thermal design has preferred side to be in sun
  - Beta (β) angle between orbit plane and sun drifts
  - Yaw flips occur when beta angle goes through zero
- "Left" and "right" swath sides are defined relative to measurement on ground relative to nadir track and do not depend on yaw state
  - End measurement given in terms of left and right sides in data products
- H (horizontal) and V (vertical) polarizations and $+y$ and $-y$ directions in KaRIn frame do depend on yaw state
  - Many L1B product variables and calibration parameters related to physical measurement are given in terms of H, V, +y and -y



Cycle and Pass Numbering



- Calibration ("1 day" or "fast sampling") orbit:
  - Cycle numbers increase sequentially from 401 to 578
  - Cycle 401 began 2023-01-15 09:26:13.011 UTC
  - Transition to nominal orbit began 2023-07-11 ~03:00 UTC
  - Repeat period is 0.99349 days (not exactly 1 day)
    - Overpass times of day will drift ~9 min earlier per repeat cycle
  - Calibration orbit has 14 revolutions or 2*14 passes numbered from 001 to 028

- Nominal ("21 day" or "science") orbit:
  - Cycle numbers increase sequentially from 001
  - Cycle 001 began 2023-07-21 05:33:45.768 UTC
  - Repeat period is 20.86455 days (not exactly 21 days)
    - Overpass times of day will drift ~3 hrs earlier per repeat cycle
  - Nominal orbit has 292 revolutions or 2*292 passes numbered from 001 to 584

- Both calibration and nominal orbit phases:
  - Ascending passes have odd numbers (001, 003, 005, …)
  - Descending passes have even numbers (002, 004, 006, …)
  - Pass duration is ~51 min
  - KaRIn data from drifting orbit phases is not processed

Height References

- SSH and PIXC height are referenced to ellipsoid
  - Ellipsoid parameters are given in metadata of every single KaRIn product granule
  - As of Sept 2023: WGS84

```
:ellipsoid_semi_major_axis = 6378137. ;
:ellipsoid_flattening = 0.00335281066474748 ;
```

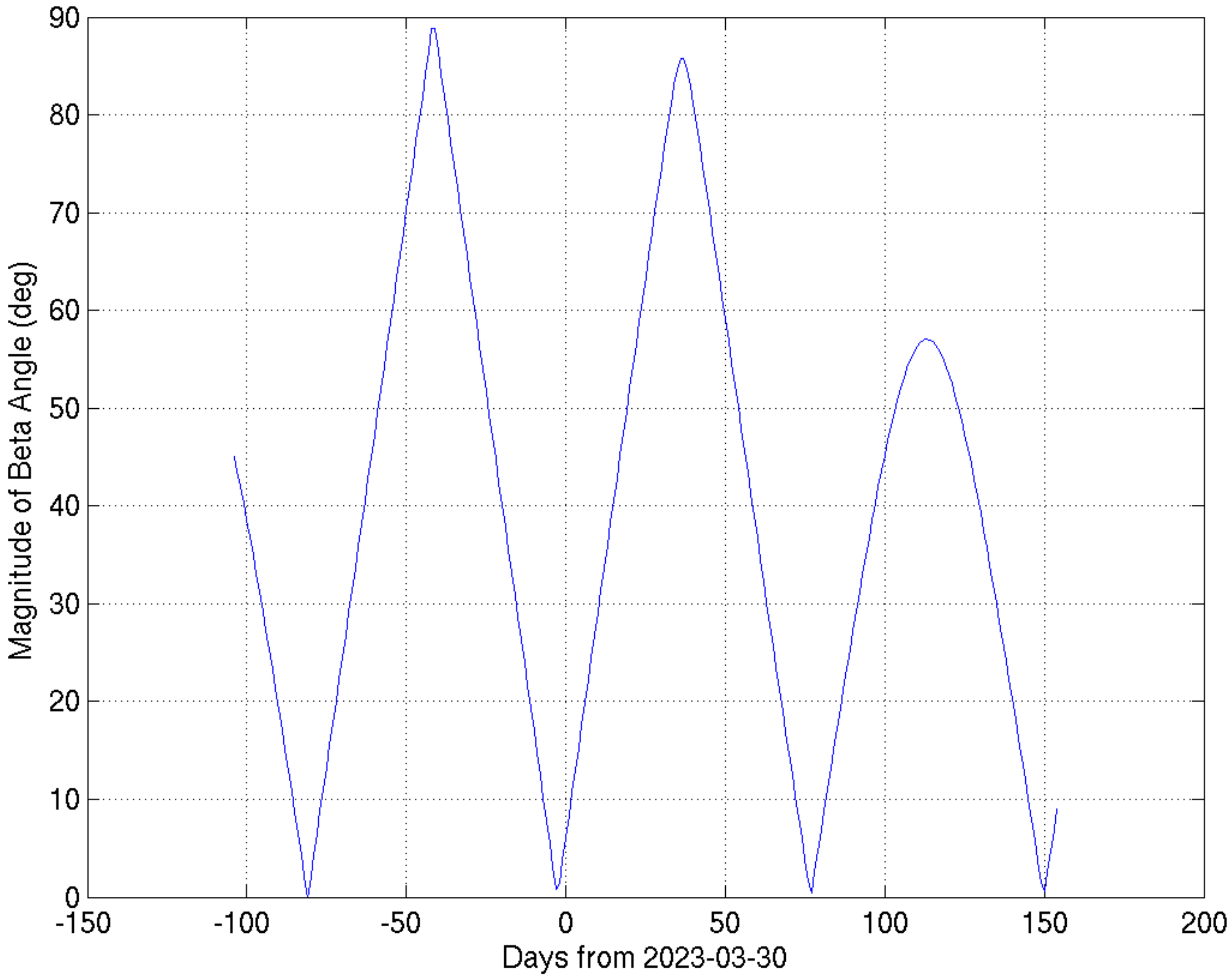
- River, lake, and raster water surface elevation (WSE) are referenced to geoid
  - Geoid heights relative to ellipsoid are reported in L2 products at each sample location
  - As of Sept 2023: EGM2008
- SSHA is referenced to mean sea surface (MSS)
  - MSS heights relative to ellipsoid are reported in L2 LR products at each sample location
  - Two MSS models (CNES/CLS and DTU) reported in product, but only one used to compute SSHA from SSH
  - As of Sept 2023, SSHA assumes
    - CNES/CLS 2015 MSS for beta-pre-validated
    - CNES/CLS 2022 MSS for pre-validated

Truth Definitions

- Truth definitions for SWOT validation may not always agree with specific definitions of quantities used by individual users for particular purposes
  – Definitions are matter of convention for SWOT
  – Consistency in interpretation is most important
  – **Users should be aware of differences in truth definitions and measurement conventions, especially when attempting to "validate" SWOT**
- Example: For validating river products, true or ideal reach slope is defined: *reach_slope = (WSE_at_reach_start – WSE_at_reach_end) / reach_length*
  – Reach length is static and comes from prior river database (SWORD), not from SWOT observation
  – Estimate of reach slope uses measurement data from entire reach to estimate WSE at each end of reach
  – Definition is equivalent to unweighted average of slope over entire reach

Relevant Time Scales of Variations

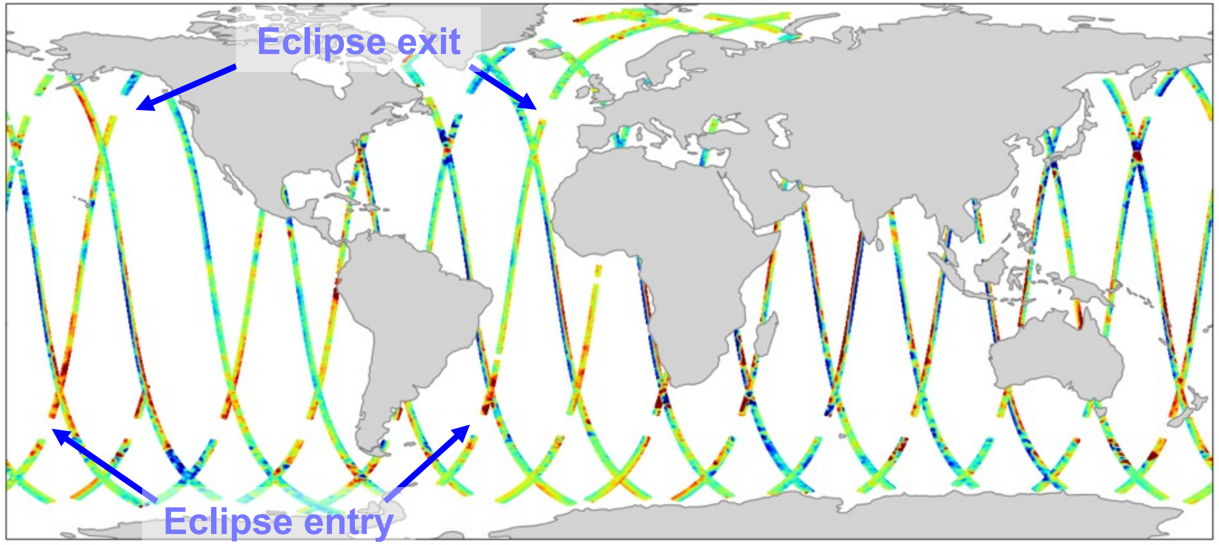
- Users should be aware of time scales of variations that may affect KaRIn data quality
  - Seconds to minutes: KaRIn parameter changes
    - Parameter changes should be compensated by KaRIn internal calibration and processing
  - ~100 min: KaRIn orbit
    - Variations over orbit should be compensated by crossover calibration
  - ~80 days: Beta (half) cycle between yaw flips
    - Changes in KaRIn and spacecraft thermal characteristics may give uncompensated errors

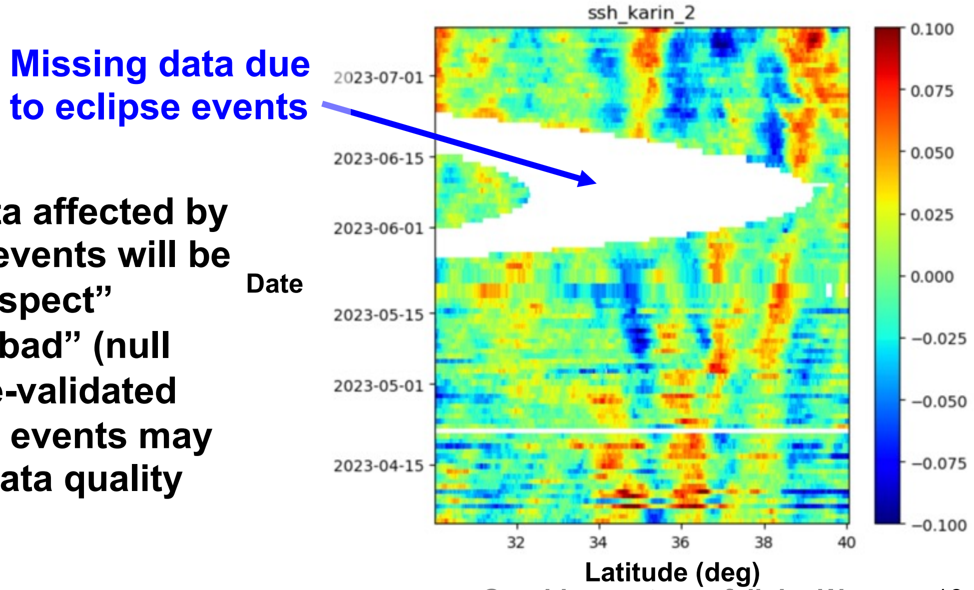


Spacecraft Events and Data Availability

- Spacecraft events impact availability of KaRIn data
  - Eclipse entry/exit: SWOT spacecraft goes in and out of Earth shadow and experiences thermal transients that can affect KaRIn stability
    - Occurs twice per orbit in similar geographic locations
    - Affects ~2 min of data (~800 km along track) after event
  - Propulsive maneuvers (station keeping, collision avoidance): SWOT needs to fire thrusters to make minor changes to orbit; involves large attitude variations
    - Occurs every several weeks
    - Affects few hours of data after event
  - Yaw flip: 180 rotation in yaw
    - Occurs every ~2.5 months
    - Affects few hours of data after event
  - Solar array rotations: SWOT solar arrays are re-oriented to collect sunlight with changing beta angle
    - Occurs several times every ~2.5 months
    - Affects ~12 min of data after event
- Other data loss:
  - Various issues with storing or downlinking data

Eclipse example from 2023-04-15 LR coverage





Much of data affected by spacecraft events will be flagged "suspect" instead of "bad" (null filled) in pre-validated release, but events may still affect data quality

Graphic courtesy of Jinbo Wang

Defaulted Fields When Measurement is Bad

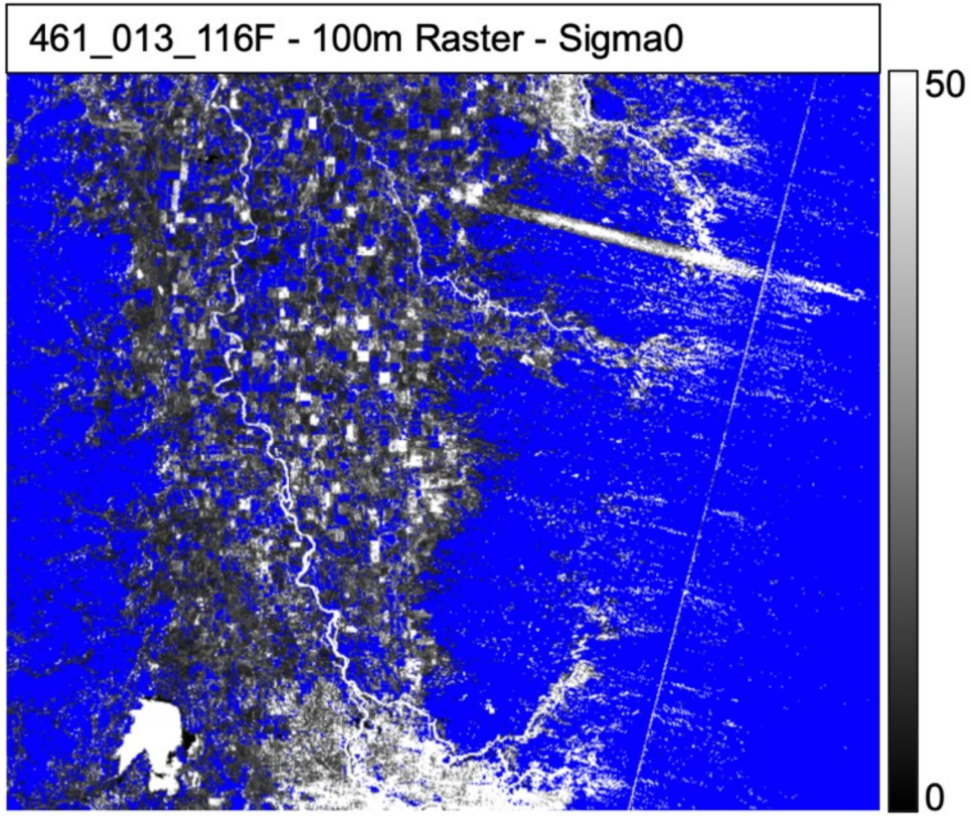
- Typical approach for averaging during ground processing:
  - L2 data products involve averaging many upstream data samples
  - Upstream data samples that are flagged as “bad” are discarded before averaging to compute downstream outputs
  - Time tags reported in L2 outputs are usually times of observation
  - Model corrections are computed at observed geolocation (measured 3-D position)
- If all upstream samples were flagged as “bad”, there are no observations to average for given L2 output sample
  - L2 output sample is flagged as bad
  - Other associated quantities may also be null filled
    - Observation time is null filled if there is no observation
    - Model corrections at geolocation are null filled if measured geolocation does not exist

No Data vs. No Detection in HR Data

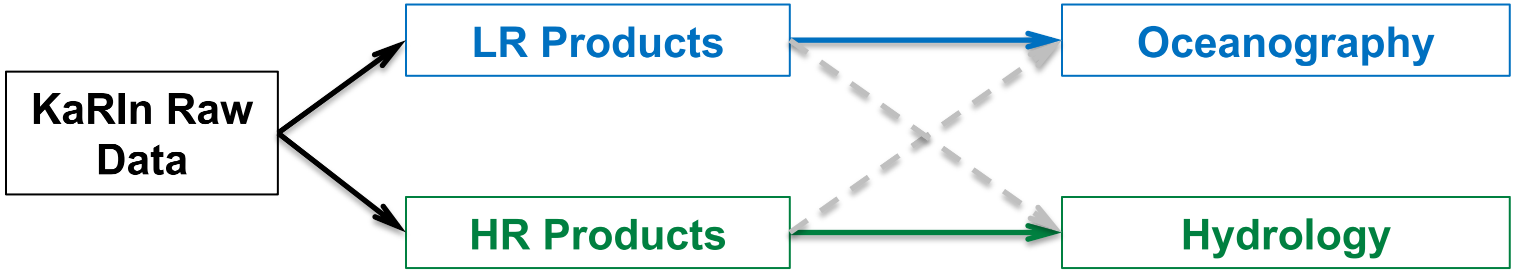
- Most land pixels are discarded from L2 HR products
  - Classification of land vs. water is done by water detection algorithm in L2_HR_PIXC processing
- If L2 HR granule does not contain any water pixels, user may want to know which case happened:
  - SWOT data were collected, but water was not detected (area was observed to by dry)
  - SWOT data were not collected (area was not observed and could be wet or dry)
- L2_HR_PIXC product contains variable `pixc_line_qual` to indicate whether data were collected
- No equivalent for LR data because water detection does not happen in LR processing

Filled River/Lake Objects and Sparse Rasters

- RiverSP product and PLD-oriented (Prior) file of LakeSP product contain one entry per database reach/node/lake that *might* be covered by granule
  - Set of reach or lake objects included in given SP continent-pass granule does not vary with cycle number
  - May includes objects up to 80 km from nadir, not just objects from 10-60 km cross track
    - Intent is to be able to report any useful observations rather than have product definition exclude good data
  - Objects are null filled if water is not detected
    - **Objects outside 10-60 km but within 0-80 km cross track may always be null filled**
- Raster product is null filled where water is not detected
  - Flags indicate approximate observation coverage to distinguish no-water vs. no-observation cases



LR Data Over Inland Water and HR Data Over Ocean



- KaRIn LR and HR data streams are split on board spacecraft in instrument firmware processing
  - Nine-beam LR interferograms are formed on board and spatially averaged before being downlinked
  - HR pulse data are pre-summed (low-pass filtered in along-track) on board before being downlinked
- Ground algorithms and data products are designed around using LR data over ocean and HR data over land, not vice versa
  - LR data for hydrology and HR data for oceanography *may* still be useful
  - But prospective users should gain familiarity with data products and algorithms to determine whether LR data for land and HR data for ocean meet their needs/desires

Users should *not* assume that LR and HR data differ only in horizontal resolution and height accuracy

LR Data Over Inland Water

- KaRIn on-board processor (OBP) uses flat reference surface per swath side
  - Design of on-board reference surface was based on hydro input
  - May still give increased error where there are rapid spatial variations in elevation
- Phase-bias correction in ground processing is sensitive to spatial variations in backscatter and topography at 1–10 km length scales
- LR ground processing does not include many steps that are done in HR processing:
  - Classification (water detection and dark water flagging)
  - Phase unwrapping with respect to HR reference DEM
  - River and lake vector processing
- Crossover-calibration corrections are not applied to LR products, so LR products will contain spatially varying cross-track tilts
  - But crossover-calibration correction terms are reported in product so users can apply themselves
- LR quality flags are designed for ocean and may not be trustworthy
- Validation of LR data over land has not been high priority to date

LR data products are *not* simply less-noisy, coarser-resolution versions of HR products

HR Data Over Ocean

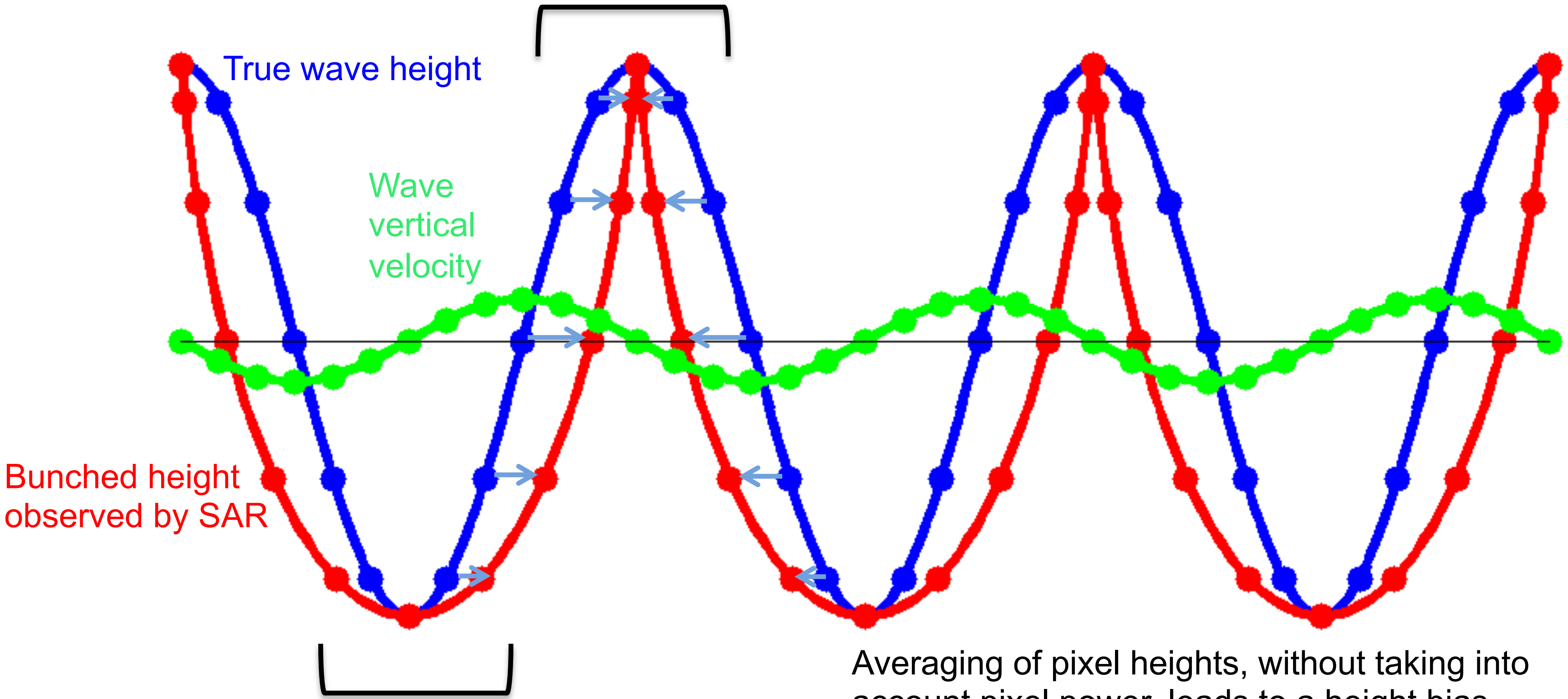
- Pre-summing in OBP implies loss of information in downlink
- HR ground processing does not do very much spatial averaging
  - HR data may be sensitive to wave-bunching effects observed on AirSWOT unless specialized post-processing is applied
- HR data products do not include ocean tide or sea-state bias (SSB) corrections
- Prior-based river and lake outputs may exist but be empty over ocean where there are no database features
- HR quality flags are designed for inland water and may not be trustworthy
- Validation of HR data over ocean has not been high priority to date

- Note: Cal/Val team uses HR data over ocean, but only for specific calibration purposes and only after customized offline processing (not available generally)

HR data products are *not* simply noisier, finer-resolution versions of LR products

Height Distortion From Wave Bunching

This pixel has a higher density of mapped points.



This pixel has a lower density of mapped points.

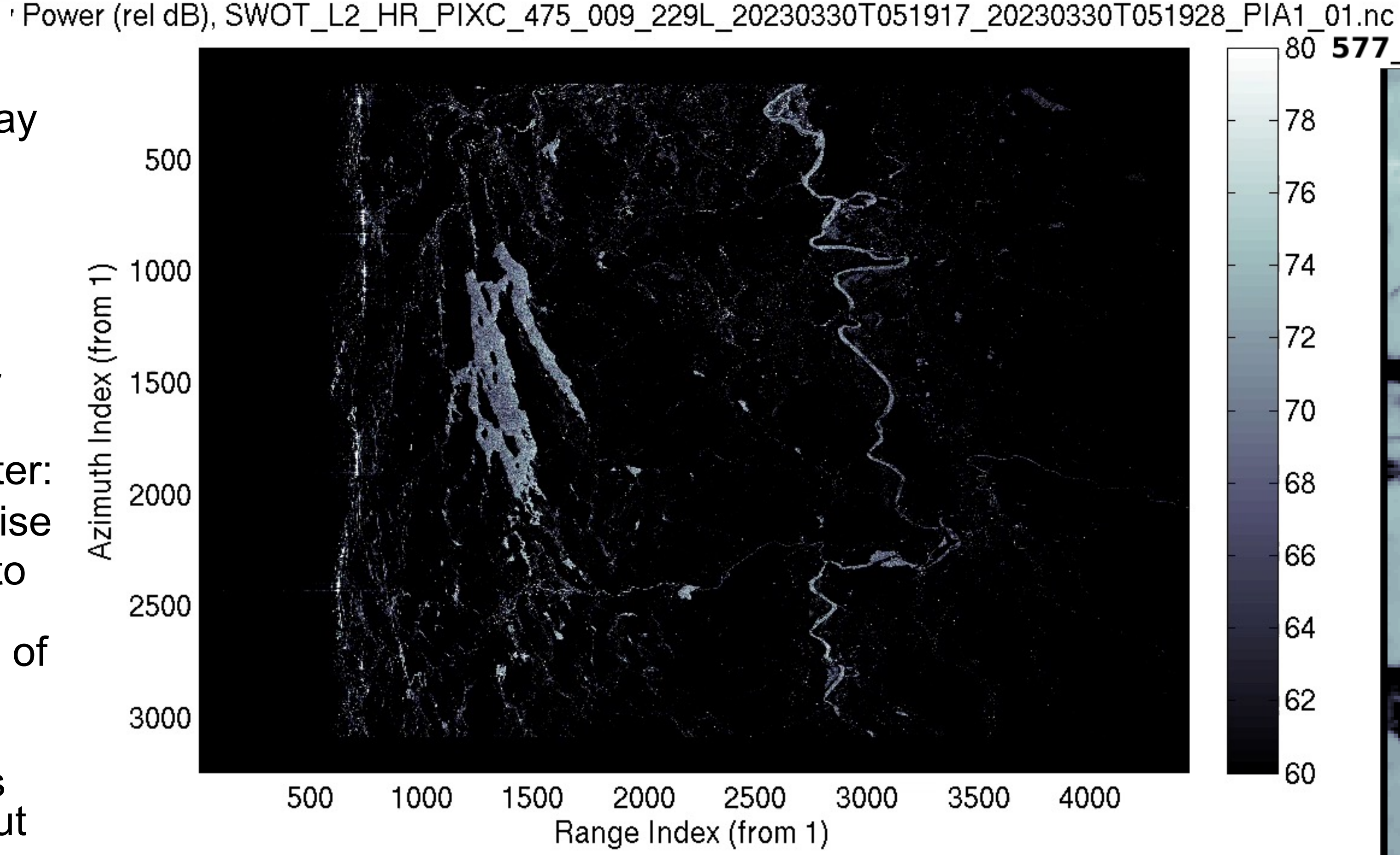
Azimuth shift is proportional to line-of-sight target velocity, which is mainly due to wave vertical velocity for near-nadir viewing geometry

Averaging of pixel heights, without taking into account pixel power, leads to a height bias. In the simple sinusoid case shown, heights would be biased low.

Wave bunching is non-linear distortion, so spectrum of observed heights can exhibit energy at spatial frequencies that are not present in the true wave field

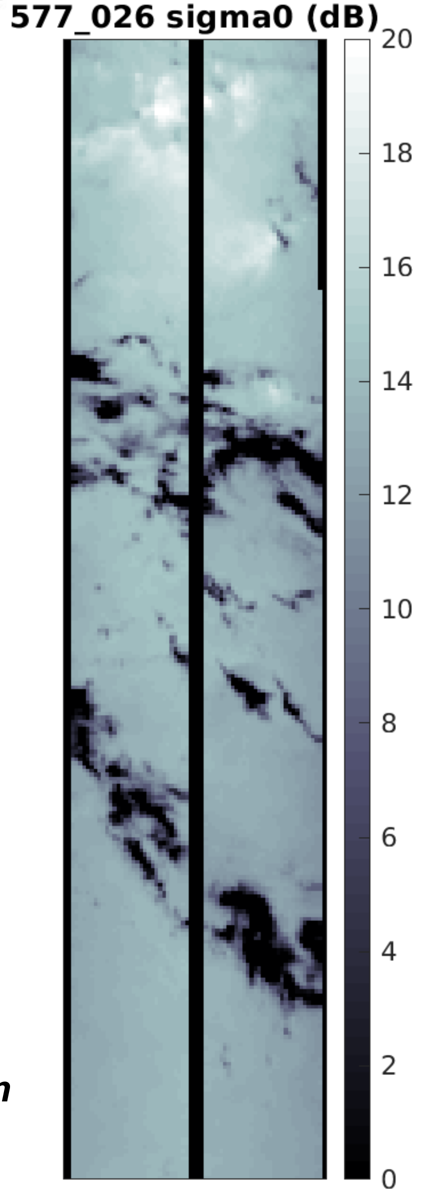
Dark Water

- Backscatter of water may be dark for different reasons:
  - Rain
  - Highly specular reflections (more significant issue for inland water)
- Effects of low backscatter:
  - Greater random noise
  - Greater sensitivity to systematic errors from contamination of nearby targets
  - Dark water not directly detected as water in HR data but may be flagged as dark water based on prior (not SWOT) data



Dark water over Connecticut River over April 2023

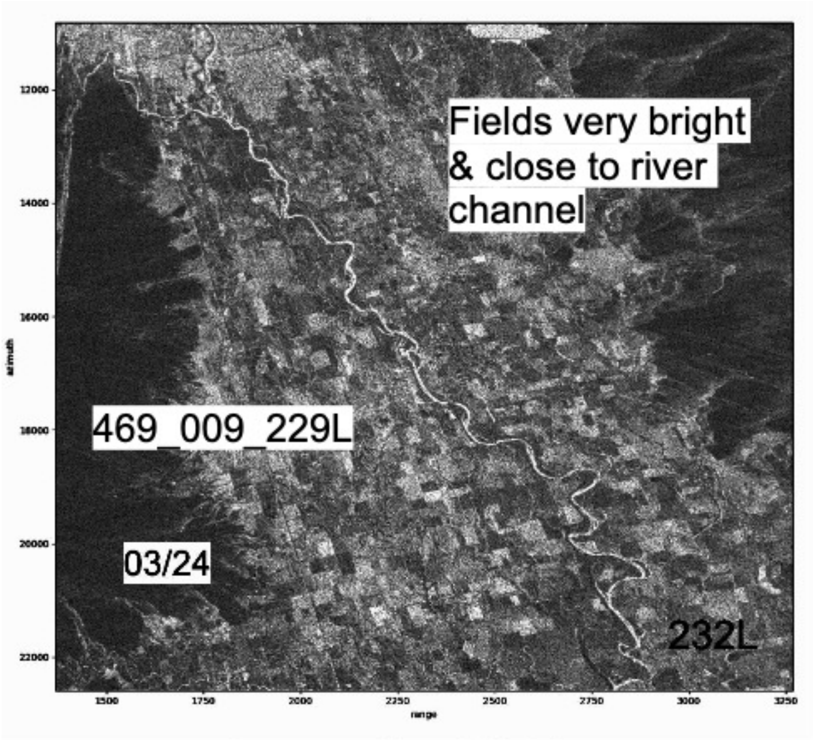
North



Effect of rain on backscatter over ocean

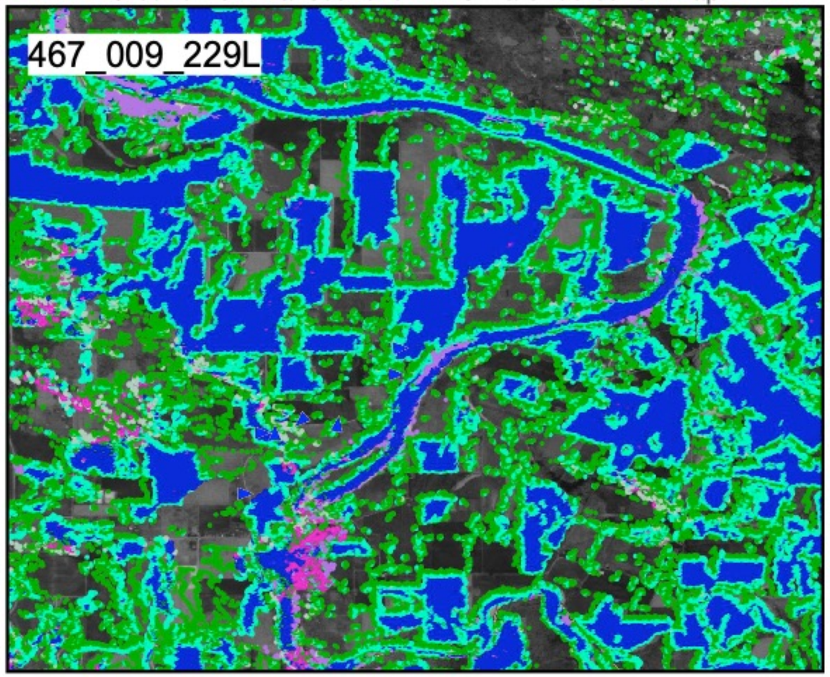
Willamette River Bright Fields Example

- Agricultural fields near Willamette River Cal/Val site are bright and are incorrectly detected as water, especially in winter/early spring data
  - Overdetection of water affects pixel mapping to river and therefore corrupts river height, slope, and area estimates
  - Fields in other areas are not as bright
  - Willamette fields became less bright going into summer
  - Overall impact to river height, slope, and area estimates needs further assessment after additional calibration work, comparison to field data, and algorithm tuning



Contrast-adjusted SLC image

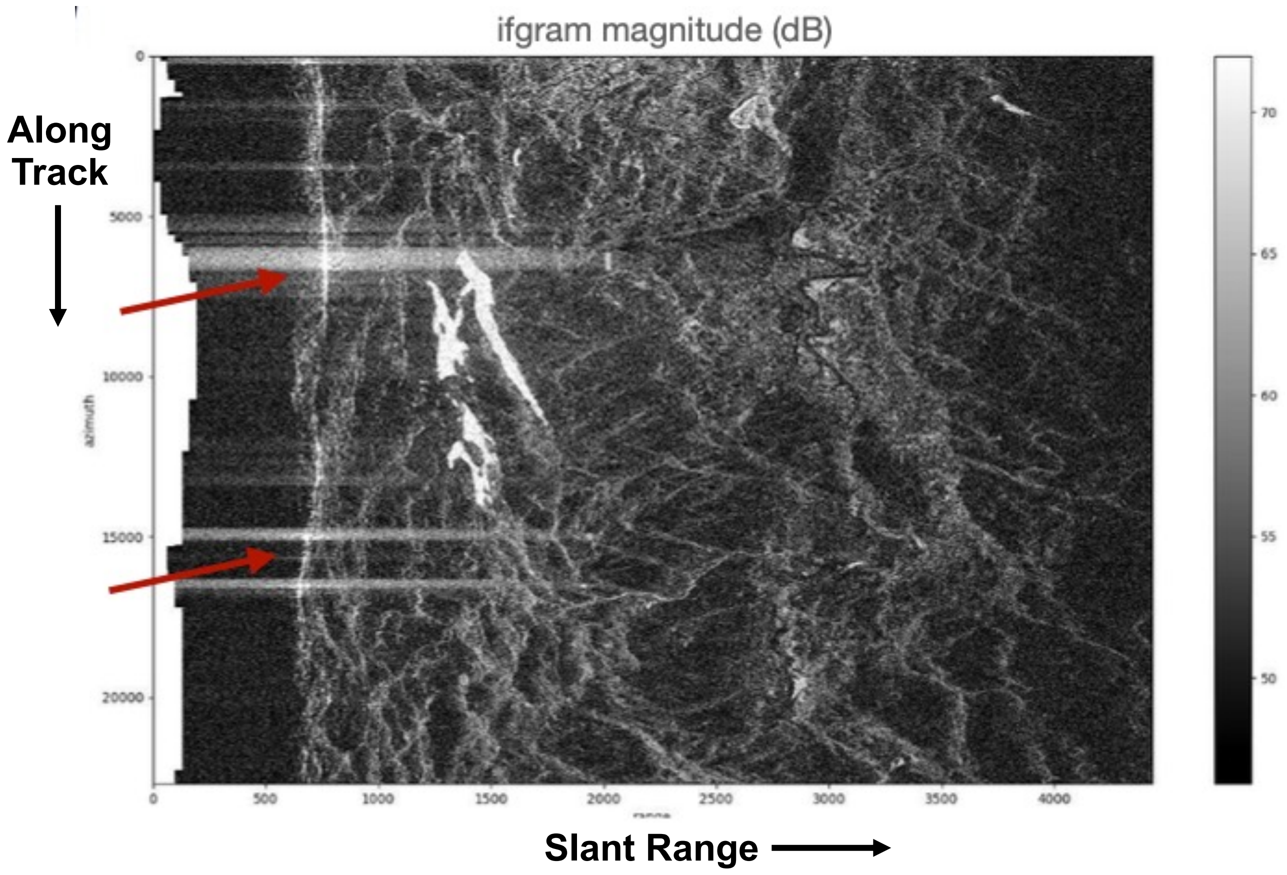
Ground truth transect over PIXC water detection map



blue=interior water; cyan=water-near-land;
pink=layover class (i.e. low coh)
purple=dark water

Bright Nadir Returns

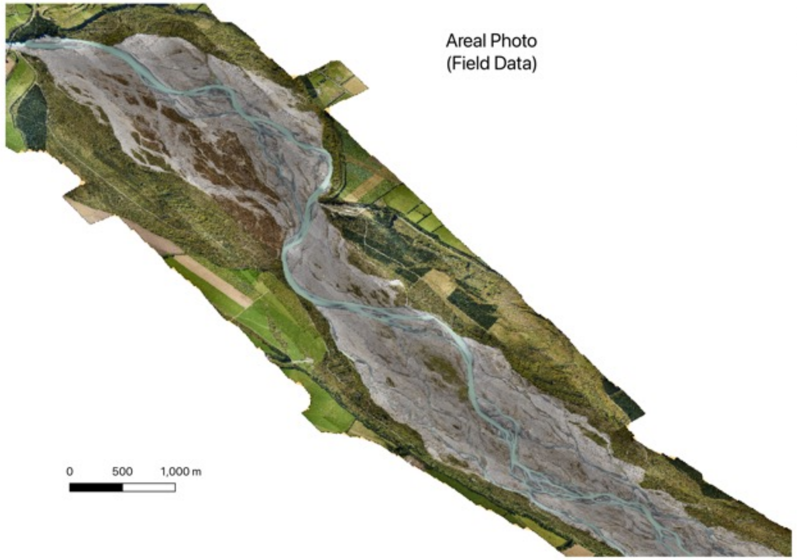
- Specular echoes from nadir are sometimes so bright that range sidelobes of point-target response corrupt other parts of images
- Algorithms have been updated to flag and ignore corrupted pixels
  - But not in beta-pre-validated release

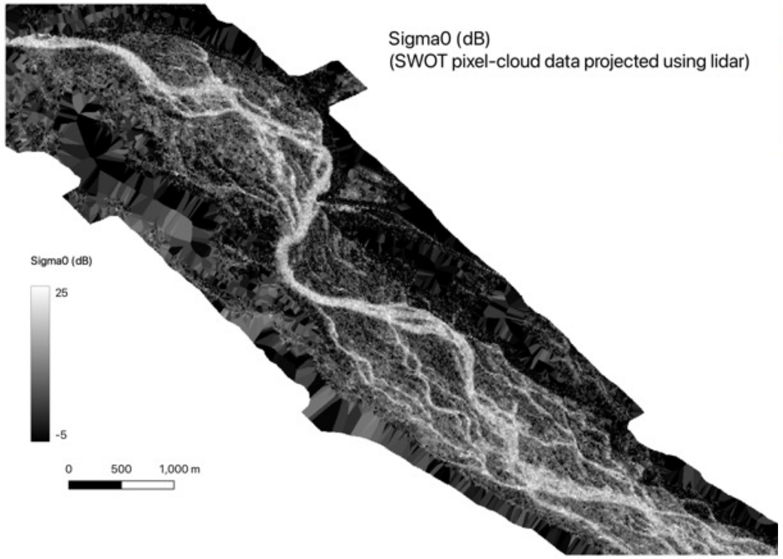


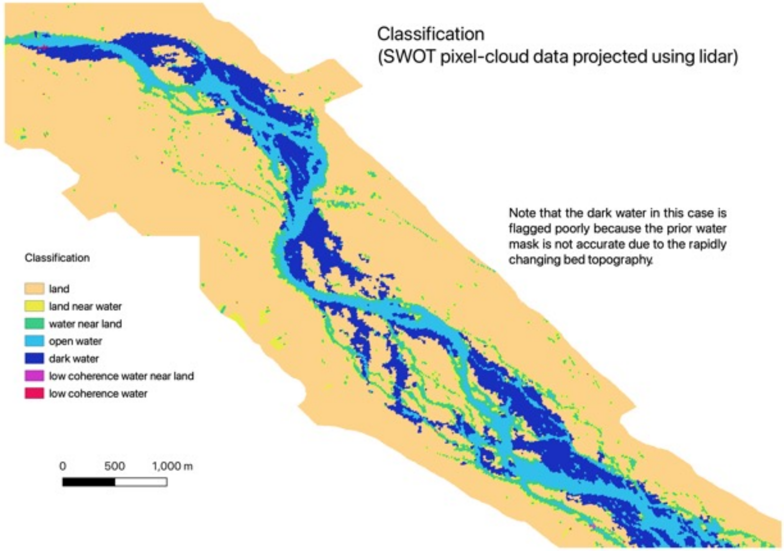
Connecticut River example with bright nadir echoes (and dark water on river)

Waimakariri River Example

- Detection of real water works quite well over Waimak
- Dark water flagging does not work well because rapid migration of river channels causes smearing in prior water probability map



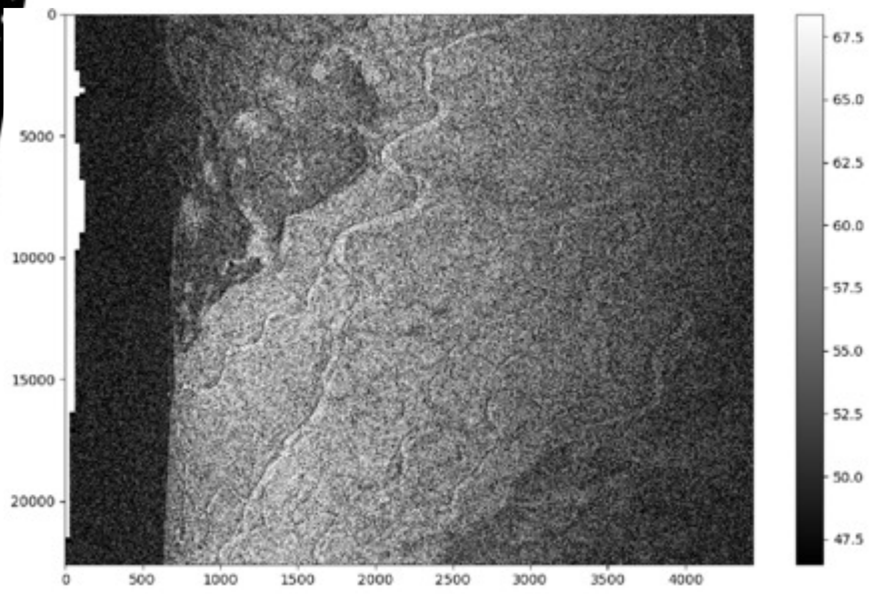




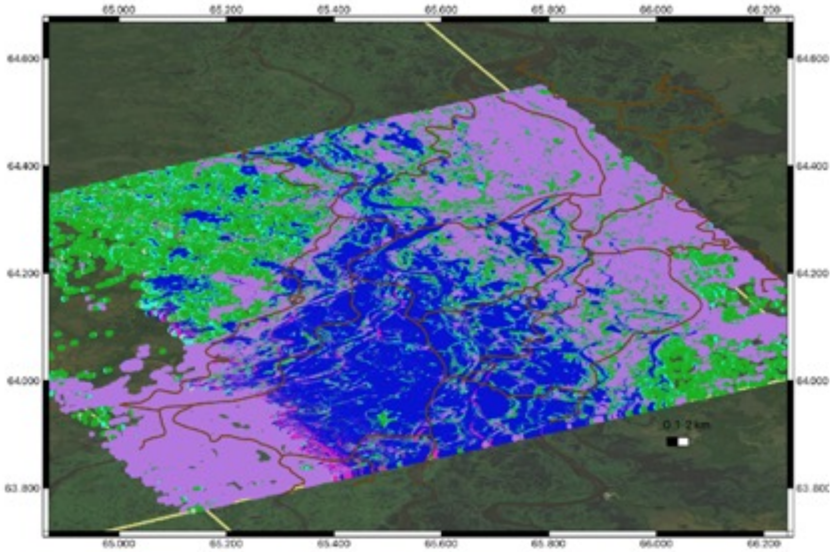
Ice

- Ice can appear bright and be detected as water, but resulting height measurements may not be reliable
- KaRIn data may have potential for cryosphere science, but validation of performance is not primary priority for current project work
- Ice flags exist in both LR and HR data products

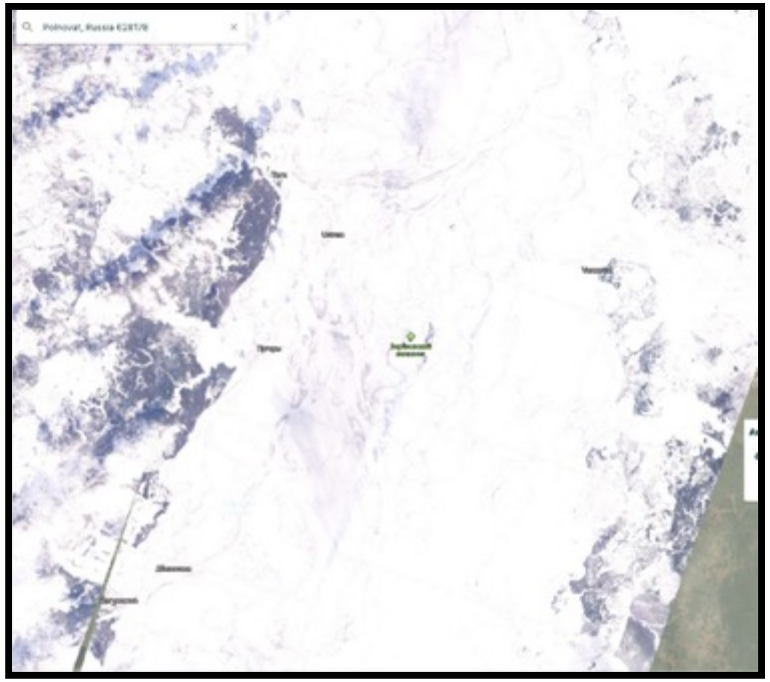
Backscatter



Classification



Planet RGB Optical Image

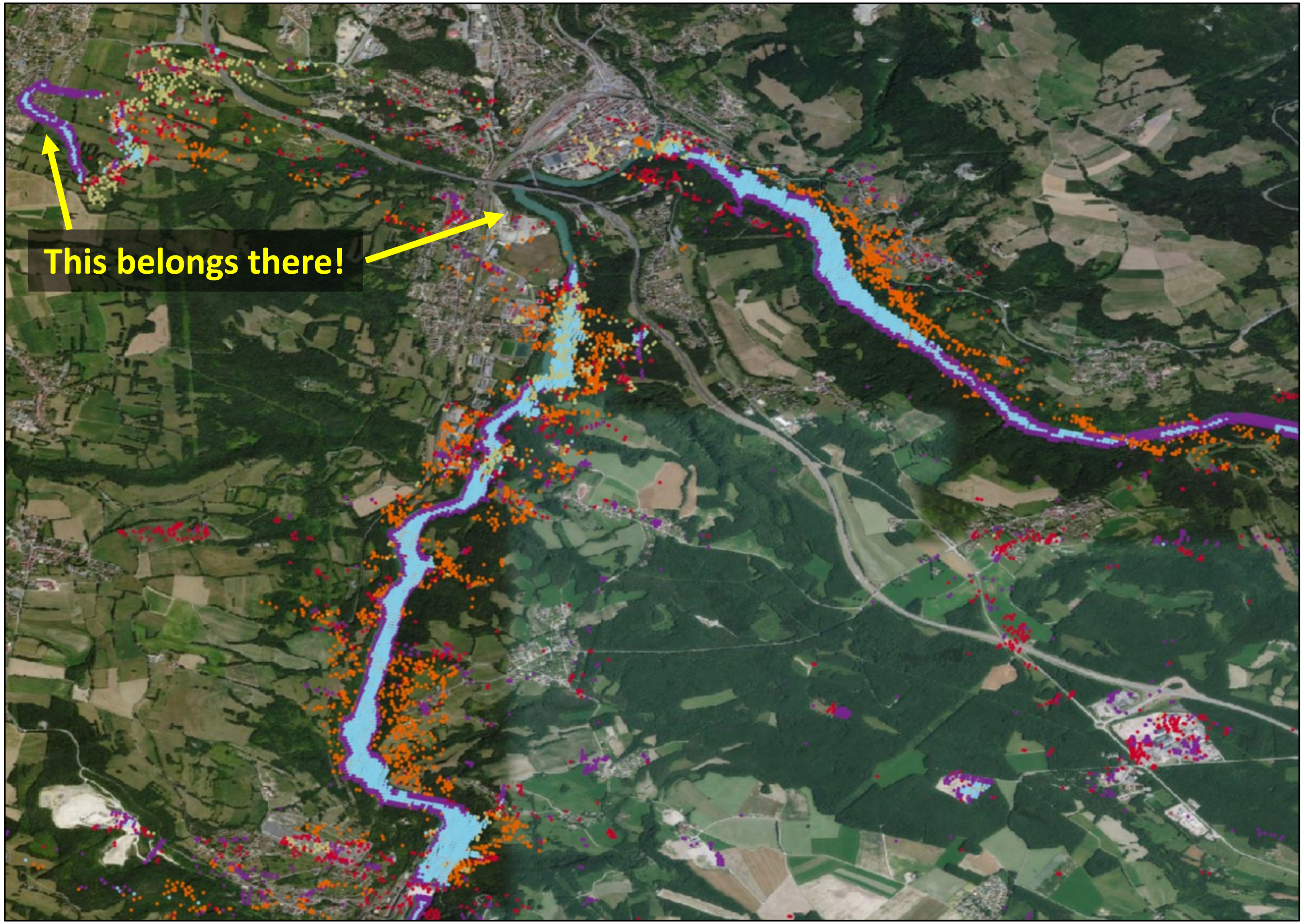


Ob River, 406_010_040L
Note: Data is from before antenna alignment and processing is not calibrated

Phase Unwrapping Errors

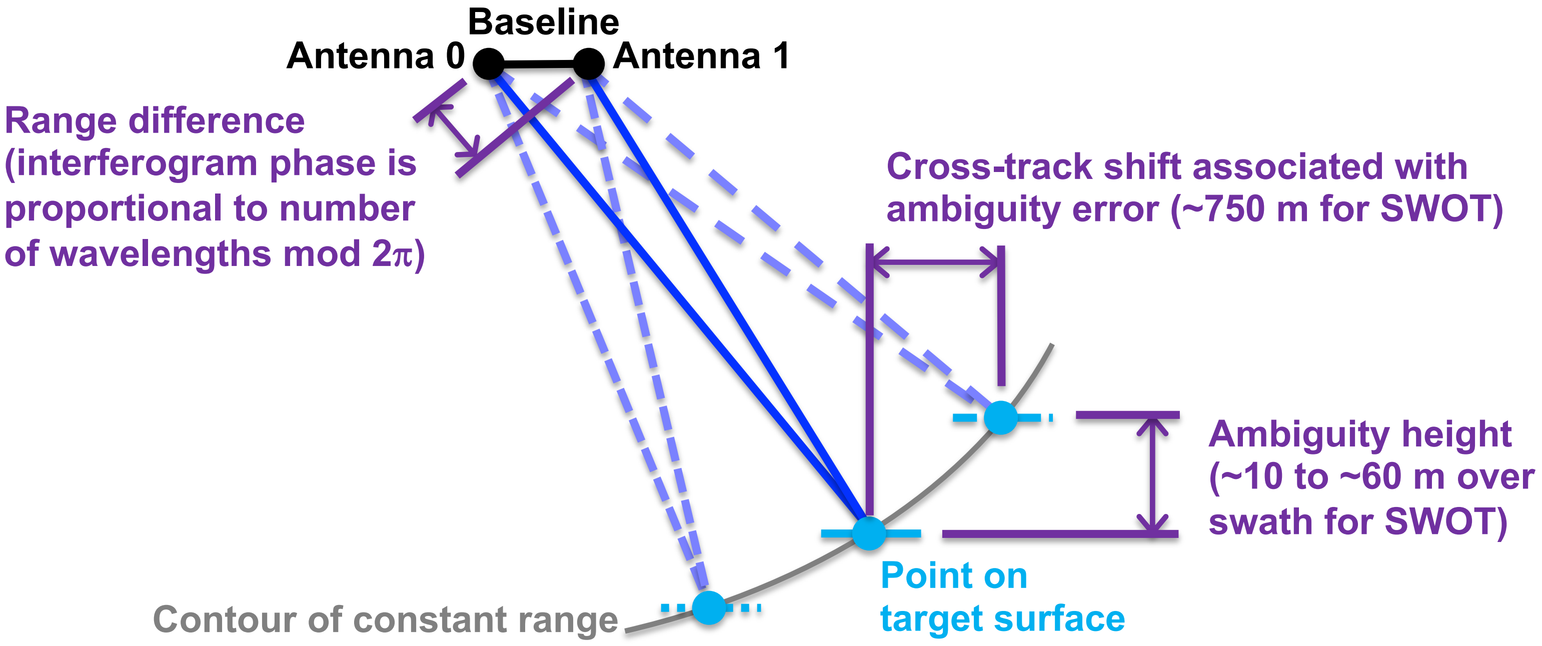
Image shows HR pixel geolocations projected onto ground and overlaid on optical layer for illustration; colors represent pixel classification values

Phase unwrapping errors can cause large cross-track shifts, large height errors, and noticeable cross-track slope errors in HR data



Rhone River, reach 21602600391, 523_003_235L

Phase Unwrapping



- Interferometric phase is precise measure of difference in range between point on ground and two radar antennas separated by known baseline
- Phase can only be determined modulo 2π radians
  - Multiple points in space have same range and interferometric phase; target location is ambiguous
  - Target location is geolocated incorrectly if incorrect phase ambiguity is assumed

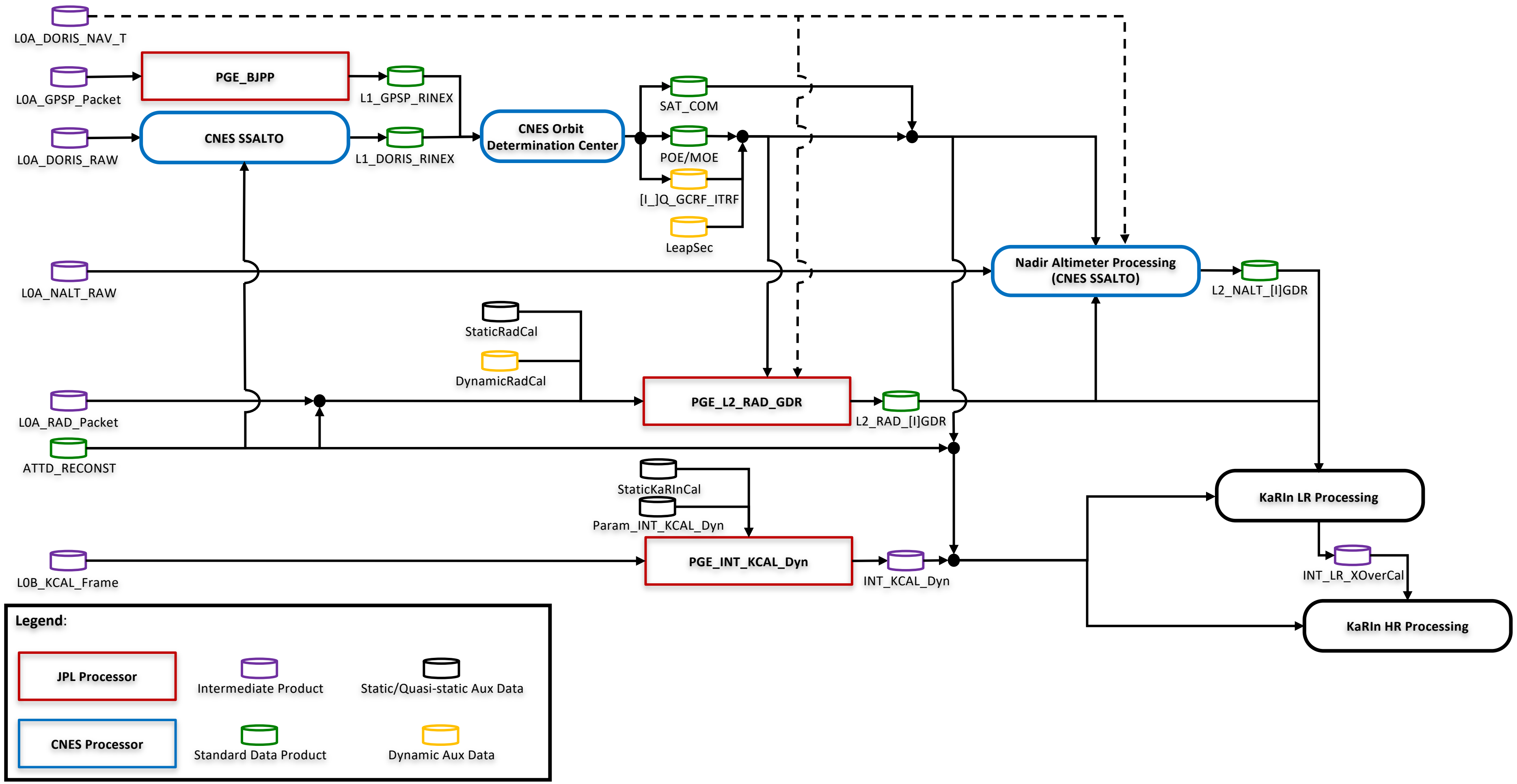
Conclusions

- KaRIn measurement is complicated
- Users seeking to validate SWOT measurements may benefit from seeking to understand measurement and conventions
  - If measurement process is viewed as black box, validation feedback may not be very helpful to project
  - Specific definitions and conventions can have significant implications on validation results
- Many sources of additional details are available:
  - Metadata of product files
  - Product description documents (PDDs) and algorithm theoretical basis documents (ATBDs)
    - https://podaac.jpl.nasa.gov/swot?tab=datasets

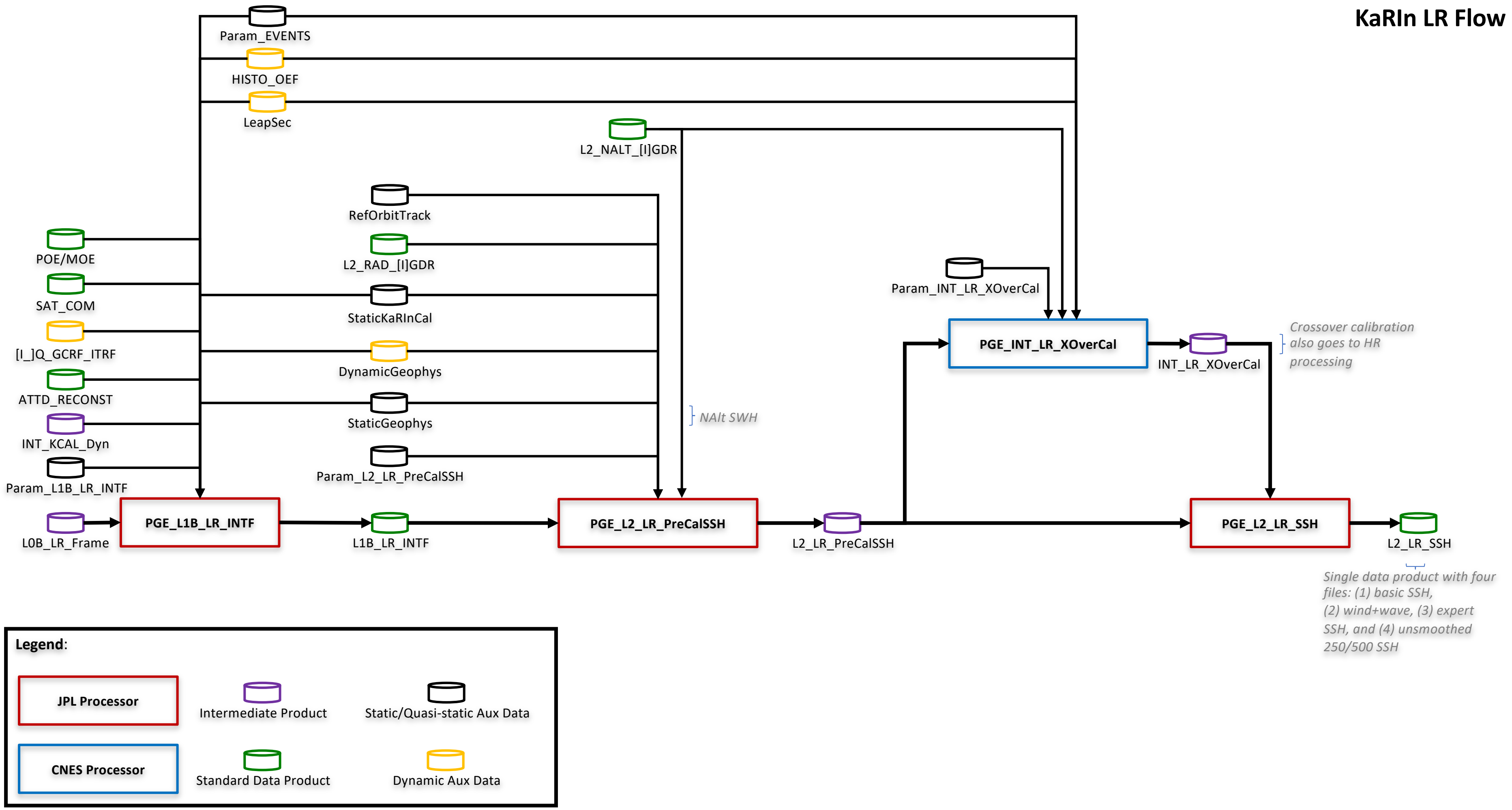
Backup

Top-Level Algorithm Flow

Radiometer, POE/MOE, KaRIn Calibration Flow



LR Algorithm Flow



HR Algorithm Flow