EOSense

# STATISTICAL DEGRADATION MODEL FOR OPTICAL SENSORS: Product Validation Plan

Project No. 18

EOSense-018-004

V2.0, 31/05/2020

Author: S. Mackin

# Table of Contents

# 1. Introduction

## 1.1 Purpose

In this document we outline the validation methods to be applied to the output from the five algorithms examined. All five are discussed, but with most focus on the three most advanced algorithms, relative gain, non-linearity and SNR.

- Evaluating the drift of on-board diffusers used in the calibration process;
- Modelling of the calibration drift without on-board devices;
- **Determination of the relative gain - the detector to detector variations in response that can produce vertical striping in pushbroom imagers;**
- **Determination of the non-linearity of the detectors in orbit;**
- **The derivation of in-orbit SNR estimates over the radiance range of the images being collected.**

Note that spectral radiance values given in this document are stated as Watts (W) but refer to ($Wm^{-2}sr^{-1}\mu m^{-1}$)

## 1.2 Structure

The document is arranged in three parts. First, the overall validation approaches are summarised. Then, for each of the algorithms identified above, an outline of algorithm operation is given, along with the specific validation approach. The final parts contains an outline of approaches to determine the uncertainty of the final data product and its impact on the validation process and some reflections from the conclusion of the project.

Throughout the document, key information on the validation process is highlighted in bold, to separate it from contextual information.

# 2. Validation Approaches

The ideal method for the validation of the methods discussed in this document would be to use the on-board calibration devices. However, the findings of the study have been illuminating, in that the on-board devices either register no problem in cases where we can clearly find evidence for a problem, or they give values that are systematically different from the values we obtain, which then raises the question of which is right.

In the first case, where the on-board device does not register the presence of the problem, which is the case with relative gain and non-linearity, we can quite simply validate the problem using visual analysis of the data. In the end the final arbiter of whether there is a visible issue or not will be the end user.

In the second case, we have to find enough alternative forms of information that derive the result in different ways that we can have more confidence in either the on-board device or more confidence in the methods we propose. This applies to the SNR estimation.

In all cases we are not saying that the method we are applying is perfectly correct, we can merely state that it is more correct and give direction on estimating the associated uncertainties.

The validation approaches for determining the absolute calibration drift without on-board devices and that of determining the drift of on-board calibration devices have been specified in this document, but at the present time rely heavily on results from a previous European Space Agency study on the AATSR instrument, as there is insufficient data at this time to finalise the best approach.

# 3. Algorithms and Validation Plan

There are five algorithms to be evaluated as listed in the introduction. They are addressed in the same order as shown. However, for the purposes of the overall study only three are fully operational and providing consistent results we can use to say something about the behaviour of the OLCI instrument (relative gain and non-linearity, and the Signal to Noise Ratio (SNR) for both OLCI and SLSTR).

## 3.1. Evaluating the drift of on-board diffusers used in the calibration process

The initially stated methodology is shown below (in the box).

This applies to both OLCI and SLSTR. The algorithm will compare, using a simple ratio, the bias subtracted diffuser value divided by the bias subtracted Level 0 value of the average across all detectors for each orbit. This will be carried out for each of the five cameras independently for OLCI and for each spectral band.

The methodology to be applied to SLSTR will be finalised after an initial analysis of the instrument design and operation but will follow a similar procedure of using the VISCAL diffuser value for an orbit after bias subtraction, divided by the average Level 0 signal value for the same orbit. For SLSTR this will be carried out for each reflectance spectral band (not the thermal bands).

The orbit data will be aggregated to daily data and the daily data finally aggregated into monthly aggregates. The validation process will vary according to the instrument being evaluated. Uncertainty estimates based on sub-samples of data, compared to using the whole data set will be examined.

This initial text (in the box above) was written early in the project and had assumptions on data availability. Since that time, with more experience of the data, a more effective way of determining the drift is suggested below.

The reality as of May 2020, is that we have a limited amount of OLCI data (5 months) that we can aggregate by date of collection and has been aggregated into weekly averages. These weekly averages seem to show quite consistent trends which although more prone to variation due to cloud cover, give a good idea of trends and discontinuities in the data. Only the first 20 bands of OLCI can be bias subtracted (using information from calibration data files). Band 21 requires a different approach which includes temperature in its formulation.

Therefore the modified approach for OLCI which should continue after the end of the project would be to continue collecting the values, bias subtract and then compare the results when they begin to overlap after the end of one year. At this point it should be apparent if there is a difference in the 2019 trend compared to the 2020 trend. Once measured, the corresponding calibration files from the two different periods can be compared and changes in the calibration coefficients and the drift can be directly compared. Any discrepancy between the coefficients and the data would suggest a drift of the diffuser, with the explicit assumption that the earth's albedo has not changed significantly from one year to the next.

For SLSTR, due to an initial set of anomalies in the SLSTR code, we only reached the beginning of the data collection for this sensor. The validation approach should also be somewhat different. In this case it is paramount not only to collect the data, but also guarantee that the overpass over the South Pole where the VISCAL diffuser is exposed to the sun is collected and processed by Eumetsat. These two sets of data, the randomly collected images, and the diffuser collected data at the pole should then be both bias subtracted using the blackbody values. The data can then be aggregated at the weekly level and the ratio of the averaged bias subtracted data and averaged bias subtract VISCAL data would be directly estimated for each week.

This ratio can then become the trend line for the following year. So for example, if we generated weekly ratio values for May 2020, then for May 2021 we could compare the ratio values and see how much they have changed. The difference, assuming that the earth's albedo was unchanging would be due to the diffuser path changing, as the electronics, detectors and optical paths would all be shared and so any change to those elements would cancel out. Only if either the diffuser path or the earth albedo changes would be discern a difference in that ratio.

### 3.1.1. OLCI – Validation

We start with an assumption, that the calibration coefficients used in the study are based on the primary diffuser alone. This should be the case, as they are the raw values derived directly from the 46 micro-band calibration images collected, these will need to be aggregated to the 21 bands of image data present in the Level 0 image product.

The drift estimate is based on comparing the change in the multiplicative calibration coefficient values between two dates against the actual change in bias subtracted counts for the same two dates. In theory if the diffuser is actually capturing the change in responsivity due to the optical path and electronics, then the change in the coefficients will match the change in the bias subtracted level 0 values. However, what happens if the diffuser itself is drifting (assuming the earth is not) then we will see a disparity.

- **Validation process – The advantage of using OLCI in this study is that there is an infrequently used secondary diffuser, which is ageing more slowly than the primary diffuser. By comparing the results over a one-year period, of the differences between the secondary and primary diffuser, we can determine how well the bias subtracted level 0 data has captured any observed differences.**

  **The difference observed between the primary diffuser data calibration coefficients and the bias subtracted Level 0 data should match that between the primary diffuser and secondary diffuser. This would give confidence in our ability to use this measure to map the diffuser degradation.**

### 3.1.2. SLSTR – Validation

The methodology employed would be similar to OLCI, we would collect bias subtracted level 0 SLSTR data, using the blackbody reference as the bias level (as was done with AATSR in a previous study). We also need to collect for each day of data, the corresponding VISCAL measurement over the South Pole. This would also need to be bias subtracted. The bias subtracted level 0 data can be aggregated at the weekly level. The corresponding bias subtracted VISCAL data can also be aggregated at the weekly level and the ratio value of the two weekly amounts determined.

Given the lack of a secondary diffuser, the usual method employed by RAL (who built the instrument) is to use vicarious calibration sites to validate the calibration performance of the on-board system. The output from RAL is a drift correction table which is provided to the end-user to correct for the on-board diffuser drift. For AATSR we compare the results from the RAL drift table against those derived from the Level 0 analysis to get some estimate of any differences between the level 0 result and the level 1 result from RAL. (The results proved remarkably similar for AATSR given that the approaches are very different apart from the fundamental use of the same data from both the on-board diffuser and from the imager.)

The question arises, what if we see “significant” differences between the level 0 analysis and the RAL level 1 analysis. I place significant in quotes, as I have not defined the level of statistical significance, which will be discussed later in this document. A significant difference might show large differences in the retrieved drift patterns, or a large percentage difference (> 1% for example) between the results from RAL and the Level 0 analysis we propose.

- **Validation process – Comparison to RAL drift results derived using Level 1 data and if disparities are noted, use additional sensors over the PICS to compare Level 1 data from RAL against more stable sensors. In this case, the approach would involve correcting the Level 1 data using the drift tables they suggest, comparing the results for a specific band (to OLCI for example), noting the differences between OLCI and SLSTR for the same sites, taking into account view angle and solar angle variations and comparing the residuals from this process to the residuals between the Level 0 and level 1 analyses. This assumes a good and stable calibration with no drift for OLCI. This will indicate the shape of any drift profile, but the absolute values will depend largely on how well both OLCI and SLSTR are calibrated after taking into account the differences in bandwidth and spectral response functions.**

## 3.2. Modelling of the calibration drift without on-board devices

This applies to both OLCI and SLSTR. The algorithm is based upon having a perfectly defined annual response function for each spectral band. By this we mean we have monitored (for example at the aggregate level of weekly data) the mean digital number level after bias subtraction of all five cameras and all bands of OLCI and all the reflectance channels of SLSTR, and for each sensor and spectral band we have perfect seasonal variation plots assuming no drift in the sensor response. Currently we have the behaviour for OLCI Level 0 data for five months as shown in figure 1.

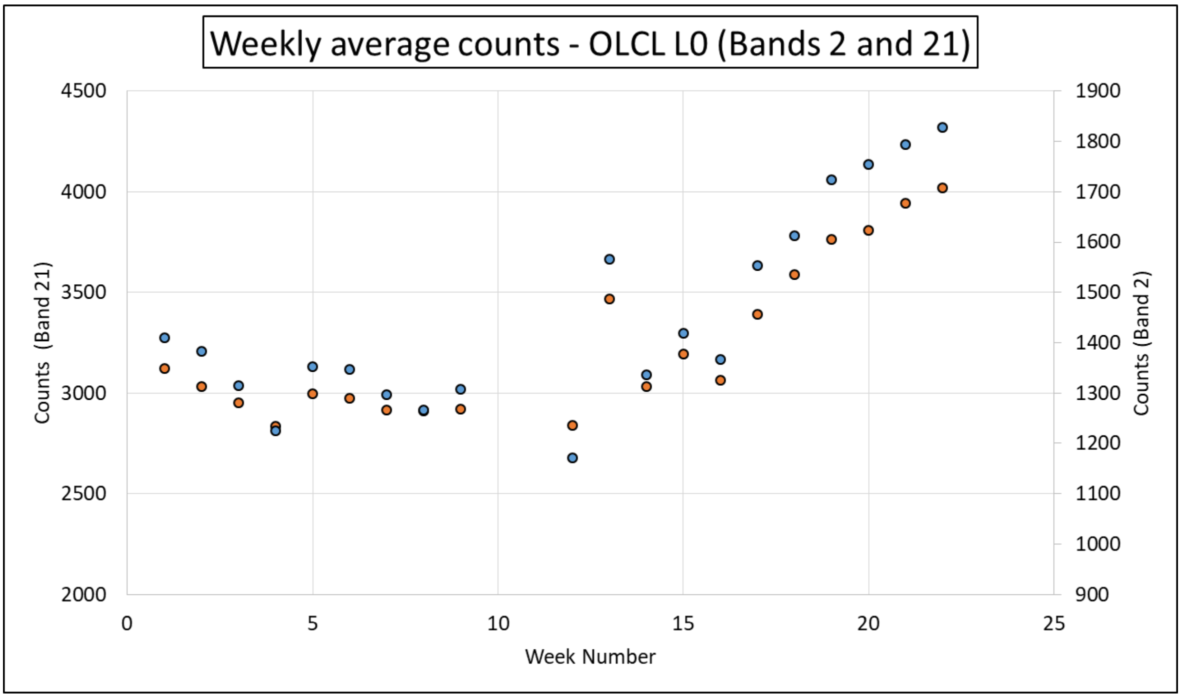


Figure 1: Weekly average bias subtracted Level 0 counts for two spectral bands.

In Figure 1, we see the normal variation in the digital number response based on weekly average values, which includes seasonal effects due to changes in the surface cover and also earth-sun distance effects. As can be seen there is a trend, the low point in this plot is around the third week of January 2020. There are also deviations from the trend, the December data seems a little higher and the January data shows a high value towards the end of January 2020. These deviations are most likely caused by for example higher than average values, such as cloud cover in the sampled data.

However, over a year we expect a well-defined cycle can be determined and therefore closely modelled, which if the earth's albedo is more or less constant from one year to the next, will be highly repeatable, assuming no instrument degradation.

We need to be aware of any instrumental changes (gain changes for example) that may cause multiplicative or additive offsets which would be seen in the average curve. These will show as a dramatic change in the data values, a large step. However, information on applied gain changes is normally registered so these events can be excluded.

There are challenges in constructing such a response curve, the curve shape will be sensitive to the bandwidth and spectral response function of the instrument in question, an approximate curve will produce approximate results. The instrument would need to be perfectly stable during the creation of the curve, which seems unlikely.

However, there are steps we can perform to generate a suitable curve for both OLCI and SLSTR. In the case of OLCI we can get bias subtracted Level 0 data for the first year of operation of a sensor and from this we can derive the curve after the following steps,

1. Correct for any gain changes or instrumental changes that would affect the Level 0 data magnitude.
2. Initial assumption is that the sensor has not changed, the primary diffuser monitors any drift, so we can use the primary diffuser to make a first pass correction to the curve generated.
3. We can use the secondary diffuser to modify the result from (2) above for any drift in the primary diffuser.

4. We can look at seasonal anomalies in data collected from other sensors such as CERES which might affect the average weekly value.

Obviously a years' worth of data provides our baseline. With each additional year, we will begin to reduce any residual seasonal effects, as they will average away with time. We can cross compare results between the A and B platforms to see if there are any sensor specific variations we need to account for. The results should improve year on year.

For SLSTR the problem is a little more difficult. We can generate the level 0 baseline curve for each spectral band. We can also modify it by the recorded drift. However, without a secondary diffuser and given the expected large drift in the VISCAL device we will accrue a good deal more uncertainty in the final result. Another approach we can consider, is to cross-compare similar spectral bands from OLCI and SLSTR in defining the average curve. In theory they should be very similar, viewing similar surfaces on the same regular basis, which, any deviation between the OLCI reference curve and the derived SLSTR reference curve should be indicative of changes required to the SLSTR curve to align it correctly.

The rather circular question, now comes back to how we validate the results we obtain.

### 3.2.1. OLCI Validation

Assuming we now have an initial first pass annual reference curve, this now becomes our reference standard. We can now determine any differences in our bias subtracted digital number values, compared to the reference standard values, as we move from one month to the next. There will be seasonal variations which we will need to monitor using external data sets (perhaps even GEO optical systems).

At this point we can now use our on-board calibration systems to give a drift estimate (using primary and secondary diffuser) which we can compare directly against the observed changes in the bias subtracted Level 0 data. Note the assumption of linearity within the whole system. We are assuming linear multiplicative changes. However, even without this we would expect to get a direct comparison of the drift component as observed using the digital numbers and those from the on-board diffuser that should be indicative and also provide information on additional processing steps to deal with issues on non-linearity.

- **Validation process – Get bias subtracted data corrected for instrumental changes. Compare to our reference curve (generated previously for the first year using the on-board diffusers). Note changes and compare to changes observed using the on-board diffusers (this is the circular part, though only for the first year). We would expect similar results in both magnitude and shape of the changes for later years based on the first year derived reference curve.**

### 3.2.2. SLSTR Validation

This is a more complex problem in that we have a single on-board diffuser that in past missions (AATSR) drifted considerably. However, we can in theory remove the drift of the on-board calibrator using the ratioing method to the earth reference as mentioned in Section 3.1 and provide a basis for estimating the change in the sensor response. Additionally, a valuable approach would be to consider similar spectral bands between OLCI and SLSTR and compare the annual reference curves from OLCI against equivalent SLSTR bands. In theory the shape of the annual curve and the variation in its magnitude should be very similar in both curves, with minor variations induced by slight band differences in terms of bandwidth and spectral response function. However, these differences should be systematic and repeatable from one year to the next. The differences will induce small additional uncertainty in the results.

This would require developing some model relationship or scaling (essentially normalising) between the SLSTR generated curve and the OLCI generated curve, which would at best be approximate, especially if there is non-linearity in each system. Based on observed changes we could estimate the change in response (in percentage terms) between one sensor and the other.

Given the tight constraints on the OLCI data through use of a stable pair of on-board diffusers, we would use this as a reference to determine changes in the SLSTR data.

- **Validation process – Use the diffuser corrected OLCI data to generate a comparison curve for SLSTR for similar bands via a model relationship. Use this SLSTR reference curve as our comparison curve for the SLSTR data that followed the reference year. The key element for OLCI and SLSTR is developing the initial annual reference curve using weekly average values.**

  **The validation is far from ideal, essentially we are determining SLSTR drift against OLCI, with the assumption that the OLCI diffuser corrected data is clean and without significant drift and that we can account or at least determine the influence of differences between the two sensors.**

## 3.3. Determination of the relative gain - the detector to detector variations in response that can produce vertical striping in pushbroom imagers

For the relative gain algorithm, we use Level 1 data that has been radiometrically corrected, but not geometrically corrected or resampled in any way. This methodology is applied only to the OLCI instrument.

We use Level 1 data, as in theory, if the system is perfectly calibrated and the detector equalisation has been performed, then the ratio of neighbouring pixels under exactly the same illumination conditions will be exactly one. What we are trying to determine is if there are issues with the equalisation and we have “persistent residuals” the name we give to the small detector to detector offsets that are consistently present in all images.

In theory these types of equalisation residuals should not be present. However, there are two occasions when they can appear,

- Type 1: When a detector has drifted in response between calibration cycles and is present for several days before the on-board diffuser is used to identify its presence and it is removed.
- Type 2: When the non-linearity correction for the individual detectors is not perfectly correct and thus we see features in normal images, but at the brightness of the diffuser on-board they are not visible. These can persist for the lifetime of the sensor if not detected and removed.

The nature of the two types of feature means we can quite easily validate the first type (Type 1) feature as we will see it in both images and diffuser data. However, for the second type of feature the diffuser is essentially useless for the validation process as it cannot detect them. This means to validate the presence of the Type 2 feature we need an alternative to the on-board diffusers.

### 3.3.1. OLCI Validation

The type 1 features should be detected using the regular on-board calibration with the primary diffuser. We would in theory determine the presence of detectors that have changed significantly in the data and cross-compare ratio of the updated coefficients to the previous coefficients to determine if both the position and magnitude of the features observed aligned. We have the additional complexity that based on the experience with Sentinel-2 it is likely that the non-linearity correction is not perfect, hence the magnitude of features we see will be a function of both the change that had taken place and the brightness of the target we use to determine the presence of the feature, so not a 1:1 match in magnitude, with that derived from the diffuser.

For type 2 features we need to use targets on the earth’s surface to at least view a small number of detectors in each of the five cameras. These sites can include water bodies, snow and desert sites that are relatively homogeneous, areas of cloud and rainforest areas may also provide a means to determine if the residuals stated are present, by simply ratioing column averages to highlight small differences in the column to column calibration for homogeneous surfaces.

We should try and match the average brightness related to the type 2 feature to homogeneous targets of similar brightness. Note that if there is some complex non-linearity relationship between neighbouring detectors, we may find that the feature detected using a homogeneous site of a specific brightness has a depth which does not match perfectly that derived from an average brightness from the relative gain algorithm.

The simplest cross-check to perform is a simple visual analysis over a water body for example, which is quite dark even in the visible bands, if we see a specific set of variations at well-defined detector locations, we should see the effects qualitatively in the imagery as a series of dark and bright stripes.

- **Validation process – The simplest validation that we are seeing significant features is the visual comparison of the presence of features over a homogeneous surface and the corresponding detection in the imagery using the relative gain algorithm. This at least proves that the features observed are not spurious. We can go further over an extended area of homogeneous material, by**

**carrying out column to column ratio analyses and determining if there is an overall offset between two highly correlated columns of data. This offset is a representation of the mis-calibration and using the average offset we will have our first estimate of the magnitude of the effect for a specific detector pair, which should be comparable to that determined from the automated algorithm, at this point we may be able to see small biases (errors) and also estimate the uncertainty.**

## 3.4. Determination of the non-linearity of the detectors in orbit

This is an extension of the relative gain algorithm applied to level 1 data that has been radiometrically corrected, but not geometrically corrected or resampled. Again this methodology is only applied to the OLCI instrument. In Section 3.3., the idea of persistent residuals seen in the data but not detected using the on-board diffuser was discussed in some detail. These Type 2 persistent residuals have been observed in Sentinel-2 and more recently in Sentinel 3 OLCI images. The non-linearity effects in the ratio values are caused by comparing two sensors with different residual amounts of non-linearity after correction, as shown in Figure 2.

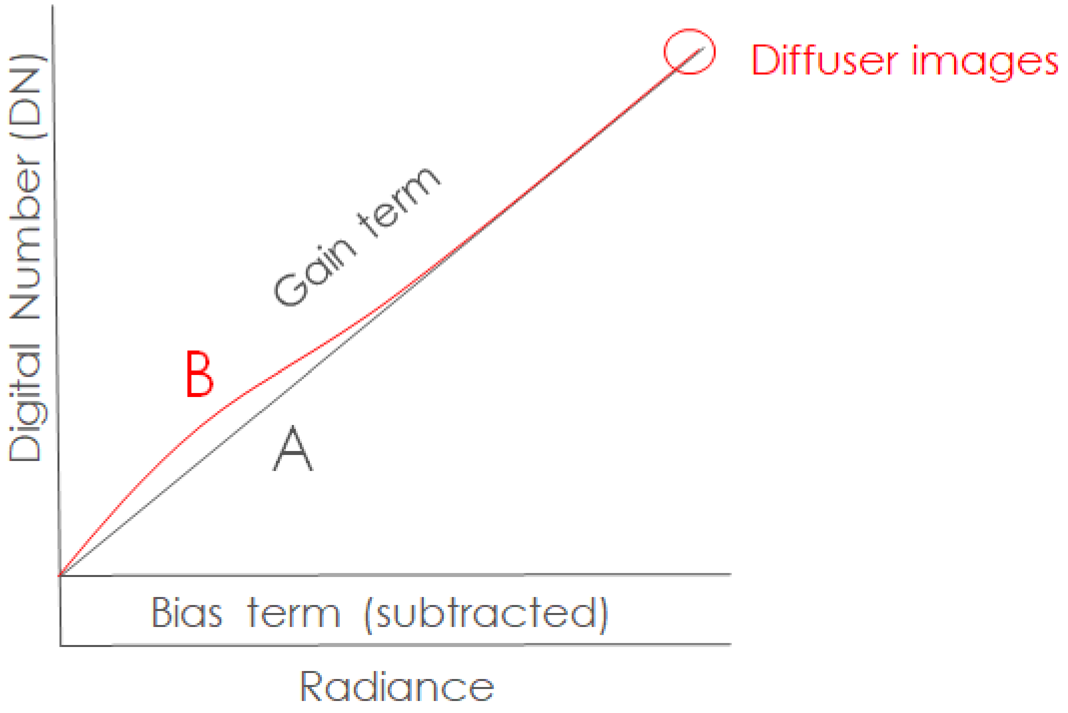


**Figure 2: Simple example showing non-linearity variations between neighbouring detectors.**

The plot shows a perfectly calibrated detector response (A) and one with some non-linear effects (B). As we ratio data from neighbouring columns, depending on the brightness, we are in different parts of the A/B ratio, from 1:1 with no noticeable residuals, to the largest offset area with features that have exceeded 0.4% in magnitude, as shown in the plot in Figure 3. Here we see Sentinel-2 data for a particular detector to detector ratio for different brightness targets, showing a distinct pattern of differences between them.

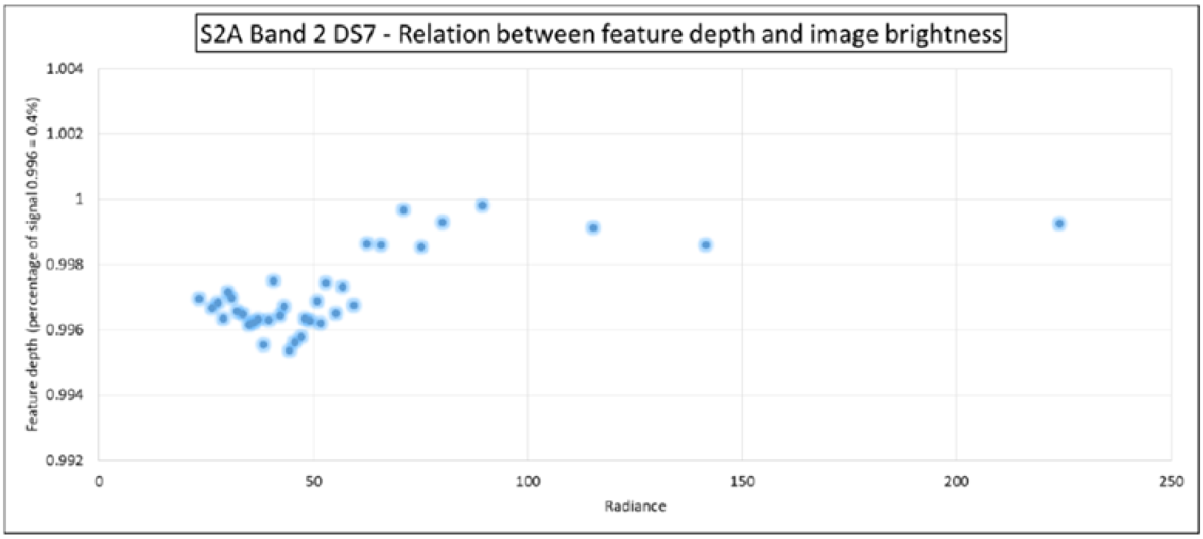


**Figure 3: Sentinel-2 data showing a relationship between persistent residual depth and surface brightness**

We need to determine these non-linear effects for each detector pair, set a reference absolute (as they are all relative measurements) and then adjust each detector to match to the reference absolute. In reality it doesn't matter in the first instance the absolute reference we use from the point of de-striping the imagery, as we are getting rid of relative differences. However, to minimise absolute errors in our estimation it is wise to get as accurate an absolute reference as possible.

In Sentinel-3 OLCI we have seen that different bands show different non-linear behaviours. For bands 20 and 21 we see strong evidence for additive term errors inducing large differences between detectors over very dark targets, as shown in figure 4.

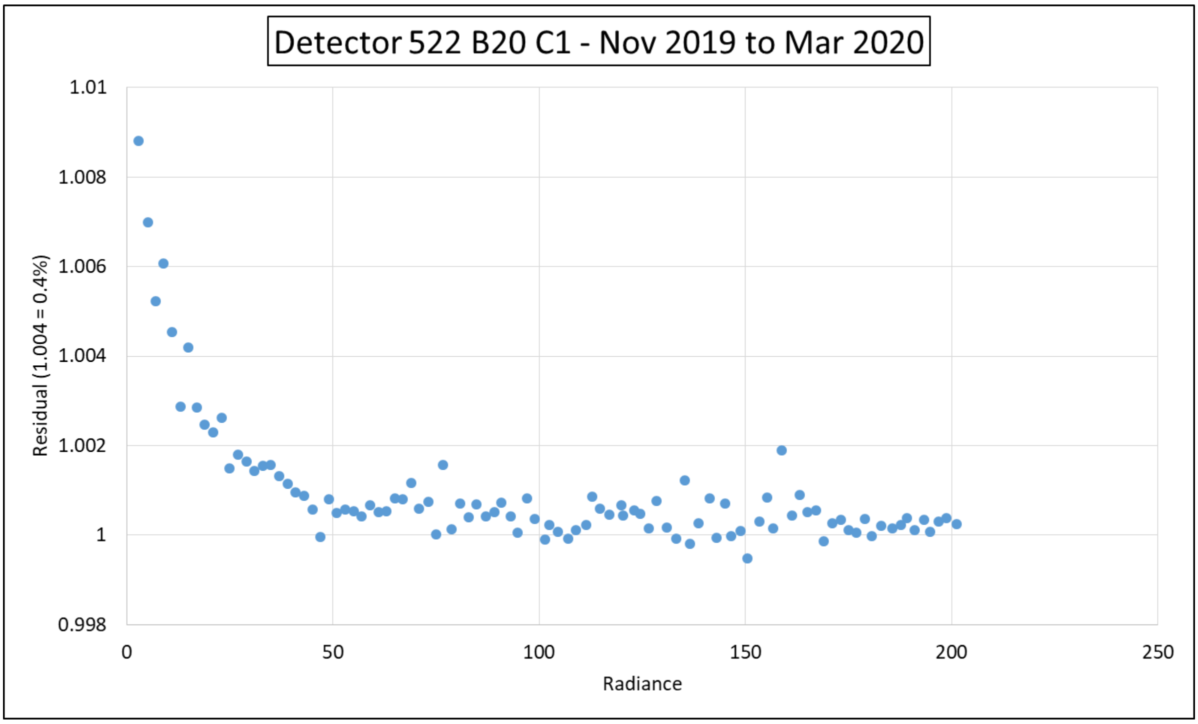


**Figure 4: Additive bias errors cause a typical non-linearity curve, converging on the "1" line.**

In figure 4 the curve approaches both axes, x and y and can be modelled by a simple additive element. A diffuser-based approach will not see this variability, although it is surprising that the dark data analysis does not record these variations.

For other bands such as Band 13, the relationship can be more complex (figure 5), with a mix of both additive terms and multiplicative terms, reminiscent of those seen in Sentinel-2.

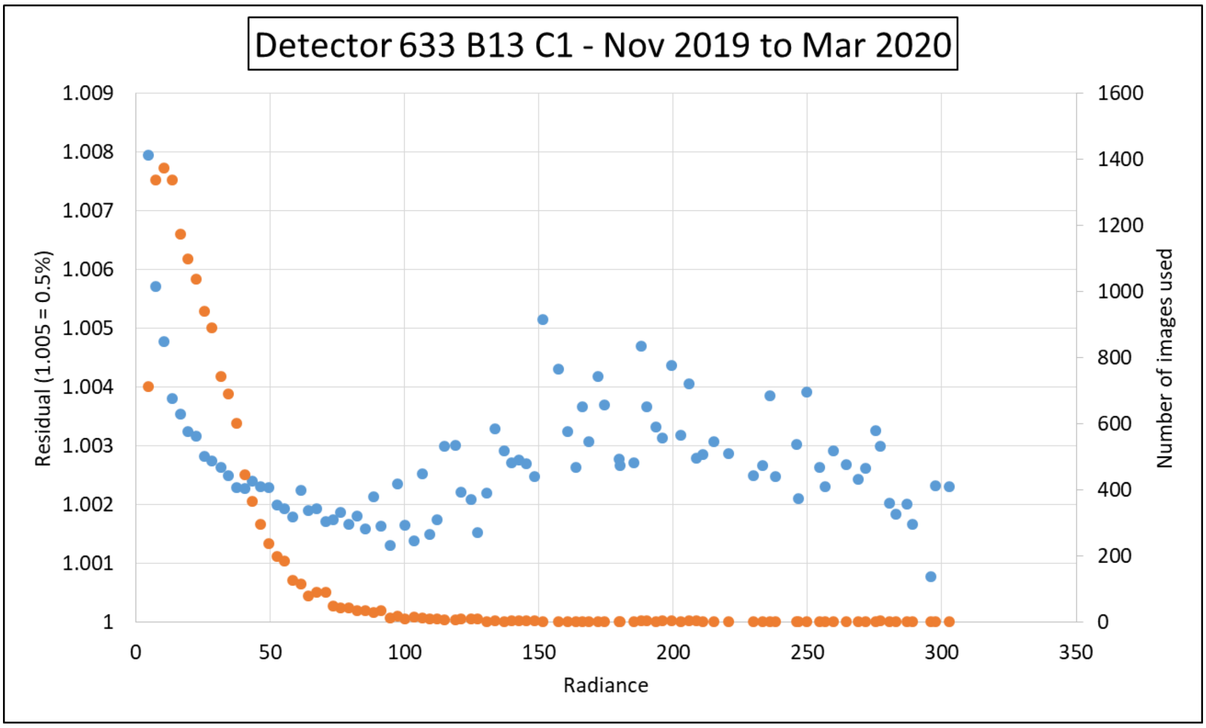


**Figure 5: Residual variations due to non-linearity (blue) and number of images processed (orange).**

The validation of such complex functions as shown in figure 5 is not trivial, except over very dark targets, where the additive term variations should be discernible. The multiplicative variations seen at higher radiances would require viewing several images to identify its presence.

### 3.4.1 OLCI Validation

The validation process in this case cannot rely at all on on-board diffusers, as they cannot detect variation in the radiance range of interest (in the example in Figure 4, below 50W). So we need to use ground targets as discussed in Section 3.3. These sites can include water bodies, snow and desert sites that are relatively homogeneous. Areas of cloud and rainforest areas may also provide a means to determine if the residuals stated are present, by simply ratioing column averages to highlight small differences in the column to column calibration for homogeneous surfaces. This is a coarse approach but simple to employ. We should try and match the average brightness related to the type 2 feature to homogeneous targets of similar brightness. Note that if there is some complex non-linearity relationship between neighbouring detectors, we may find that the feature detected using a homogeneous site of a specific brightness has a depth which does not match that derived from an average brightness from the relative gain algorithm.

If we look at Figures 3 and 5 we can see some scatter in the results, even for relatively narrow radiance ranges (as values are binned in this plot). The scatter comes from several sources - for example, if we are matching data from two neighbouring detectors in an image. Generally there is a very high correlation between neighbouring columns in an image. However, this is not always the case (for example over urban areas, or very heterogeneous scenes). In this case when we ratio neighbouring values we get a lot of variability, which makes the assessment of the relative gain change more unreliable and can lead to incorrect values for some scenes, as the average brightness might seem the same from one image to the next, but the calculated relative gain value is highly unreliable as the distribution statistics for the column to column relationship is extremely non-normal.

An example would be two columns with a very dark and very bright feature that is found in the first column and an equal number of times found in the second column, the averages of the two columns would be the same a value which is mid-way between these two extremes. However, a distribution of the ratios would be bimodal. Reality is somewhere less extreme but would induce effects depending on the surface variability. Note that the extremely bimodal effect we could actually deal with, reality however is far more complex.

An iterative approach may in the end be the best to minimise residuals, where once a set of correction coefficients is derived, we re-run with the correction and look for lower level residuals and correct them. At some point during the iteration we will reach a plateau, a value which we cannot improve on, which should indicate the limits in our ability to remove these effects and hence an estimation on the residual uncertainty.

However, in the first instance the validation will be against homogeneous or almost homogeneous test sites of known brightness, to determine if (a) the features are present with the magnitude we estimated and (b) that the depth of the features varies in the same pattern as that defined by the algorithm using multiple homogeneous sites with different brightness values.

- **Validation process – Will use multiple targets of different brightness that are as homogeneous as possible. We should be able to determine the magnitude of any residuals and their shape relative to their neighbours. This we compare directly to the output from the relative gain algorithm for specific brightness ranges. We should see a direct and clear correlation between the automated extraction and our manual extraction over specific homogeneous targets. From this we should get a good measure of any biases between the homogeneous surface results and the heterogeneous automated method as well as some indication of the uncertainty in the final result based on the scatter.**

[Since this original documentation was written we now know that persistent residuals do exist in OLCI data and that they do show a variety of non-linear behaviours depending on the spectral band. Having reviewed the data collected over a five month period, the approaches given in this section are still valid. The only issue is to find the suitable homogenous targets. However, the approach is still a valid and simple exercise, although a little time consuming.]

## 3.5. The derivation of in-orbit SNR estimates over the radiance range of the images being collected

The coarse windowed algorithm will be used for both OLCI and SLSTR reflectance channels. We have noted that quantisation of the signal level can lead to a more approximate estimate of the SNR and has caused some issues with the modelling approach. We also have very limited information on the effects of spatial resolution as most of the early development work has focused on instruments with spatial resolutions between 1m and 30m, with only a few experimental observations showing quantisation in both MERIS and AATSR data. Once we have a suitable number of test images, we should be in a better position to evaluate these two potential issues. All data should be Level 1, radiometrically corrected but not resampled in any manner.

NB: Since the above text was originally written, we now have sufficient evidence to suggest that we see no issues with the spatial resolution from OLCI or SLSTR, at least in deriving the SNR data clouds. Additionally, we see no evidence of the effects of quantisation on the results. Hence the data generated within this project for both OLCI and SLSTR seems valid at the time of writing of this plan (May 2020).

### 3.5.1 OLCI Validation

The SNR is estimated at diffuser brightness using the on-board diffuser and is presented in the cyclic reports for the sensor. The values presented in the cyclic reports seem stable, although significantly higher than those observed in the one piece of pre-launch data that we have available. The ideal validation would be against the diffuser derived values. However quite early in the project we determined that the diffuser estimates seemed much higher than those we derived, although the relative band to band variations were comparable and that our estimates were much more in line with the pre-launch estimates. Given this discrepancy we had to determine if there were alternative validation methods that did not use the on-board.

We therefore ran our process on white scenes, relatively homogeneous snow scenes from Antarctica, altogether 14 images. The estimates from these homogeneous surfaces matched our original estimations very closely and also matched the pre-launch evidence. Only the diffuser data gives a higher than expected SNR.

- **Validation process –This can only be carried out using multiple sources of information, given the findings of disparities with the on-board calibrator estimates. We propose the use of homogeneous surfaces and pre-launch data to assess the SNR.**

### 3.5.2 SLSTR Validation

We will use the blackbodies and the diffuser to estimate the noise at two specific extremes, plus estimates from the literature based on pre-launch measurements and comparison from interpretation over specific homogeneous sites.

- **Validation process – We have as yet to compare the results to those stated by RAL. No validation has been attempted at this time. Using the bright diffuser, it should be possible to at least get a single tie point to estimate the SNR and by modelling in the same manner as in the cyclic reports from OLCI then extrapolate to produce SNR curves under the assumption of shot noise limited systems. These can then be fitted to the data clouds produced by our algorithm to determine how well we capture the variation.**

# 4. Uncertainty Estimation

As part of the process of validation we have outlined in this section, the possible approaches to give an adequate measure of the uncertainty involved in our parameter estimation, to allow statistically significant tests to be performed. Below are the suggestions for each parameter.

- Evaluating the drift of on-board diffusers used in the calibration process;

The simplest approach is actually a measure of the variability within the data from multiple observations. By determining how the data is distributed, the mean value and standard deviation, given a large number of samples we can determine the standard error of the mean. This will give uncertainty bars on each point for the estimation. This would allow us quite simply to vary our sample size from one day to one week to one month and determine the residual uncertainty on the estimation and use these as uncertainty to compare against the uncertainties determined for the diffuser.

- Modelling of the calibration drift without on-board devices;

The same methodologies employed in determining the accuracy of the drift estimation could also be used to determine the uncertainty in the modelling of the calibration drift without on-board calibration devices.

- Determination of the relative gain, the detector to detector variations in response that can produce vertical striping in pushbroom imagers;

The uncertainty will apply to type 2 persistent residuals, as every image will contain them. The same method as described for calibration drift can be used, to get the standard error of the mean, but keeping in mind that we are mixing data from different target radiances, hence the variability we see will probably be greater, but it will give at least some estimate of the uncertainty, although slightly biased to be higher than it should.

- Determination of the non-linearity of the detectors in orbit;

For non-linearity we will use the same method as for relative gain, except this time we work with smaller radiance bins. The problem with the smaller bins is that when the number of data values becomes small the standard error can be quite large. By using data over bigger time periods, so instead of using weekly data we can use monthly or quarterly data, we can get a reduced variability. However, we also need to be aware that if the time interval becomes too large, then any drift in the relative gain values with time will induce increased variability in the final calculation. It may require some iteration with different time intervals and choose the minimum if it is well defined as the uncertainty to use.

- The derivation of in-orbit SNR estimates over the radiance range of the images being collected.

There are two possible approaches, the SNR estimation based on the Settle Model uses a model fit with an “error” term associated with the fit. This model uncertainty would normally be the baseline to use as we model the SNR curve over the radiance range of the sensor. However, the initial algorithm implemented used the simpler peak based statistics. To validate this data would require a larger number of images, where we replicate the measurement for approximately the same radiance values and with a large enough sample, determine the mean estimate and standard error on the mean. Obviously with the large amount of data potentially available we can derive a very good estimate with the large number of samples available.

# 5. Final Comments

At the start of the project, the validation of some of the results was considered relatively simple, as given the presence of on-board diffusers on OLCI we could validate the SNR and relative gain values with ease. However, we found that the SNR values from the diffuser were tens of percent higher than those from estimates using heterogeneous images, homogeneous snow images and from limited data we had pre-launch. The cause of the disparity between the diffuser and all other estimates of the SNR for OLCI is unknown at this time.

For relative gain, again we believed we could use the diffuser to validate any features we saw in the data. However, although we can see quite large features in some bands, the diffuser analysis does not detect them. The only conclusion is that given that OLCI only has a two-point measurement strategy, dark shutter and bright diffuser, is that if non-linearity existed between these two points then we could induce detector to detector variations that were persistent across all images and that varied in depth with brightness.

It was a setback for the validation that our primary source of validation information could not be used. Hence we have had to resort to an approach based on a more manual and time consuming analysis, including visual analysis of the data to find the presence of the features we have determined using our automated processes.

For the Level 0 analyses, the outlined methods have been given, but no attempt has been made to evaluate them in this project as we do not have a suitably complete data. However, given the success of the approach for AATSR comparing the Level 1 drift estimates against the AATSR drift estimate based on level 0 data, we fully believe that the methods will be appropriate.