EOSense

STATISTICAL DEGRADATION MODEL FOR OPTICAL SENSORS:

Validation and Evolution Report

Project No. 18

EOSense-018-039

V2.0, Date 06/06/2020

Author: S. Mackin

Contents

Executive Summary 3

1 Introduction 5

1.1. Purpose of Document 5

1.2 Scope of Document 5

1.3. Applicable Documents 5

2. Results - Level 0 Processing and Analysis 6

3. Results - Level 1 OLCI SNR Estimation 8

4. Results - Level 1 OLCI Relative Gain Estimation 19

5. Results - Level 1 OLCI Non-Linearity 31

6. Results - SLSTR Level 1 Processing and Analysis 38

7. Summary Discussion 42

7.1. Future Activities 43

7.1.1. Level 0 43

7.1.2. Level 1 43

7.1.3. Interface Elements. 44

7.1.4. Additional Validation 44

7.1.5. Application to Other sensors 45

8. Conclusions 46

Executive Summary

In this document, we discuss the results of the analysis of data collected from the OLCI and SLSTR sensors of Sentinel-3A, using novel methods of data quality extraction based on normal heterogeneous images.

The analysis is split between Level 0 data analyses, focused on absolute calibration drift and on Level 1 analyses covering Signal to Noise Ratio (SNR), Relative Gain and Non-linearity.

The Level 0 data analysis requires a large amount of data collection over an extended period for definitive validation, and although within this study we can point to consistent behaviour in weekly aggregate data from OLCI (and promising performance), we cannot at this time provide any conclusive results until a year's data has been collected. For SLSTR, due to an early anomaly with the code, we have even less data than for OLCI so we do not expect a result from the SLSTR Level 0 analysis until spring of 2021. Nevertheless the early results look very promising for the planned comparisons.

For the Level 1 data analysis very good results were obtained for both OLCI and SLSTR. For OLCI three areas were examined, SNR, Relative Gain analysis and Non-Linearity analysis. For SLSTR, only SNR was examined given the unusual SLSTR sensor design.

The SNR analyses of OLCI showed similar relative behaviour to that determined using the on-board diffuser (giving confidence in the performance of the tool), but a disparity in the absolute magnitude of the SNR recorded was noted, with the diffuser giving SNR values tens of percent higher than the heterogeneous images, limited pre-launch analysis and homogeneous snow scene analysis.

The Relative Gain analysis identified the presence of type 2 persistent residuals - features which we could identify with this novel approach, but could not be seen by the on-board diffuser. The presence of these residuals was validated by visual analysis in the imagery - striping effects which according to the on-board diffuser analysis, could not exist. The cause of the residuals and the reason why they could not be identified by the on-board diffuser seems to be related to an additional signal component, producing a non-linear behaviour.

The Non-Linearity algorithm showed that for shorter wavelength bands, we could see small calibration errors, which disappeared at higher radiance values, hence could not be seen by the diffuser. For mid-wavelength bands (bands 10 to 14) the behaviour was mixed, showing very strong features at very low radiance values due to an additive term and then similar behaviour to the shorter wavelength bands, while longer wavelength bands (Bands 18-21) were dominated by additive term errors at very low radiance values. It is suggested that some form of wavelength dependent stray light issues may be present.

The SNR analysis for SLSTR showed good results, but suggested some unusual source of noise in the imagery in some cases, producing unexpected SNR profiles.

In summary, the results have validated in part the currently applied data quality measures, but more importantly have pointed to areas not covered by current on-board devices that contribute to lower data quality, especially over ocean waters with low water leaving radiance. The methods applied indicate that there may be technical issues in assessment of SNR for both OLCI and SLSTR and that there are issues with Relative Gain in OLCI which produce significant problems at low radiances and cannot be addressed using on-board devices.

Finally there is some suggestion as to the possible source of the Relative Gain issues and a possibility of correcting the data and improving the data quality, especially water leaving radiance.

Overall this activity has proved the potential for assessing the evolution of operational optical payloads, such as OLCI and SLSTR, without using any on-board device and relying only on operational observations. In doing so, it has demonstrated its power to inform on and improve the quality of data from existing sensors and represents an exciting possibility for integration into the routine cal/val of future sensors.

This approach is entirely novel in the field of calibration and data quality assessment using statistical distributions to determine variations at a level of sensitivity usually confined to on-board instrumentation, but beyond that, illustrating some of the limitations of the current approaches and potential causes of some of the observed problems with data quality.

This new approach has highlighted the possible causes of features observed across multiple bands and illustrates that the current approaches used to reduce these types of features (relative gain variations for example) can sometimes increase the magnitude of the problem, as the problem is not fully understood. These new techniques point in some cases to the cause and hence potential solutions to the outstanding problems.

Further work is required, not just in extracting the information, but more in understanding what this extra information is telling us about the behaviour of the instrument and how we might derive a suitable methodology for removing the observed anomalies based on the physics of the issues uncovered.

<u>Notes on results.</u>

Eumetsat has processed the Level 1 data for both OLCI and SLSTR since the beginning of November 2019. There is an identified gap in the data, with data missing from the 1st to the 19th January 2020. However, there is more than sufficient data to carry out all the analyses and make firm conclusions of the effectiveness of the techniques, even with this gap in data collection.

1 Introduction

1.1. Purpose of Document

These reports cover three main elements

- A discussion based commentary on the results obtained to present.
- Evaluation of the validation tests outlined in the Product Validation Plan (PVP) [AD1], incorporated into each relevant section.
- Formulation of recommendations for future work, again largely incorporated into each section.

In detail we will discuss

- The results of the comparisons of the algorithm outputs against the validation data sets, where suitable validation data sets are available.
- Analysis and interpretation of the results and an overall assessment of the usefulness of the algorithms.
- Any open issues for further investigation (including with the validation datasets).
- Recommendations for further validation activities.
- Fixes and evolutions to the algorithms and their outputs.

1.2 Scope of Document

Note that it was agreed at a progress teleconference on 29/10/19 that

- V1.0 will be a skeleton with pre-validation for SNR and RG only, plus drafting notes to indicate possible future content
- V2.0 will be a full discussion of all validation and evolution issues.

From a validation viewpoint this document will only cover those elements in which we have a defined reference and a suitable data collection during the project to fully evaluate the algorithms applied. Hence, although there are Level 0 results in the discussion for OLCI, the results are incomplete and hence it is impossible to validate them at this time. In place of the validation discussion will be a re-iteration of potential validation approaches to examine, given the initial results obtained.

1.3. Applicable Documents

[AD1] Product Validation Plan 018-004 V2.0, 31/05/2020

[AD2] https://www.ioccg.org/groups/Inst_cal_WS/S3_S3VT_OLCI_status%20final%20small.pdf

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

2. Results - Level 0 Processing and Analysis

The Level 0 reading algorithms have been developed and fully tested. The OLCI Level 0 reader has been producing results from imagery processed by Eumetsat since November 2019. However the SLSTR Level 0 reader had issues. These issues were identified and corrected in February 2020. Given the lack of processed SLSTR Level 0 data it will not be discussed further.

The OLCI Level 0 data was condensed down to give averaged image values (excluding dark pixels) for each of the 21 bands and the smear band. These were stored in five separate camera databases. The initial analysis was to treat all the data values for a specific band as a time series to see overall trends and any discontinuities in the data (figure 1).

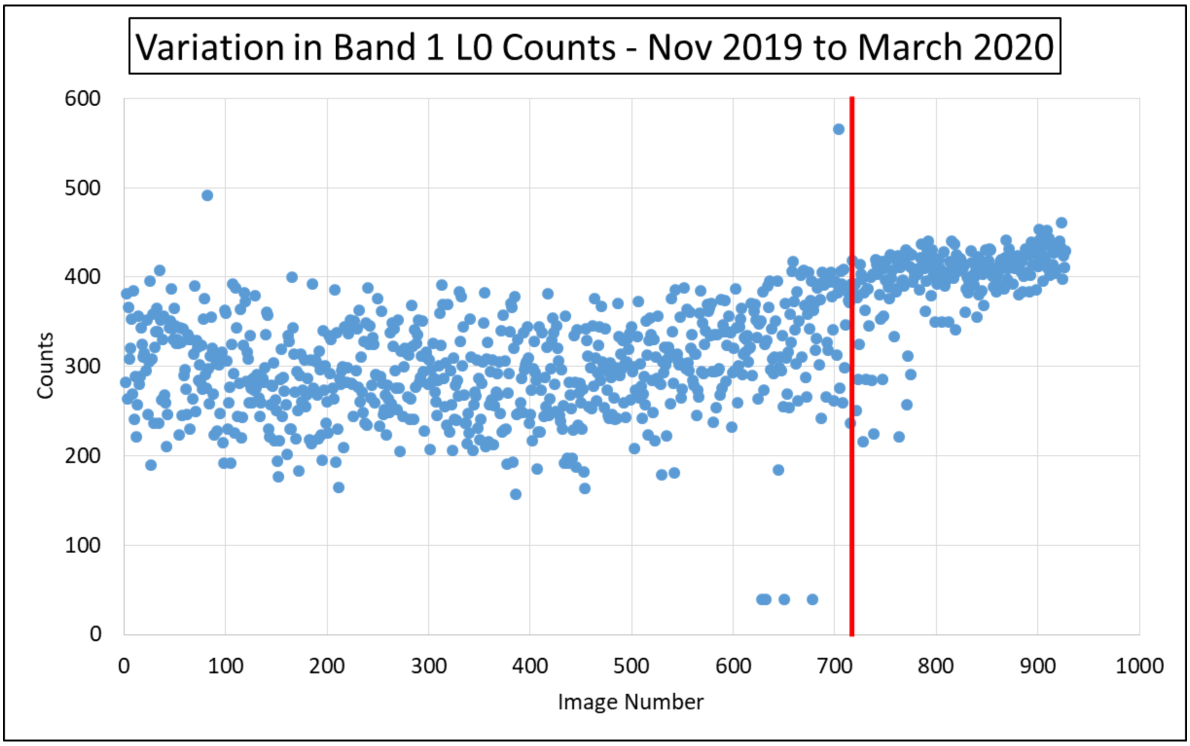

Figure 1: S3A OLCI Band 1 L0 counts variation. Red line is the 1st March 2020.

There was as expected, a large degree of scatter. However, the behaviour does seem to change as we move towards the right side of the plot, beyond the red line at 1st March 2020, with much less scatter after that date. Given the number of values involved, this change is unlikely to be a random effect, but as to what causes this reduction in scatter, it is not entirely clear at this time, but may be related to seasonal changes. The same effect was seen in multiple bands.

By opening a camera database in Microsoft Excel, we were able to collect average values for each band for that camera using the date of acquisition to select them by various date ranges. So the same data as figure 1, but averaged into specific time intervals (figure 2).

We decided that we would begin with weekly intervals. These interval averages were then plotted against week number to determine if there was any particular trend in the data values (figure 2).

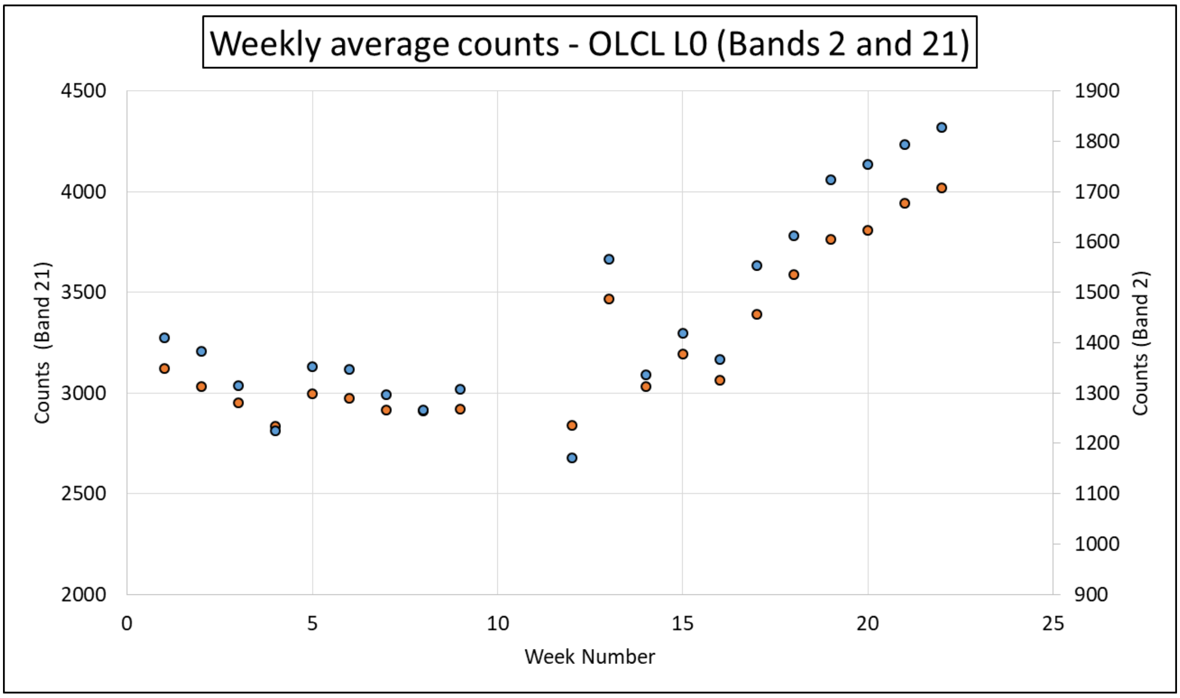

Figure 2: Weekly average OLCI L0 count values for two separate bands

The plot (figure 2) shows the results from two widely separated spectral bands, which suggests that both bands are showing similar behaviour. We at first considered that the pattern might be related to solar irradiance changes, but given that in early January (roughly week 10), the earth is in its closest approach to the sun we would expect an inverse mapping, with higher counts in January and lower in July. Hence the conclusion is that overall brightness changes, perhaps due to distribution of the continents, with more land area in the northern hemisphere, is biasing the overall results.

In winter Siberia and Canada are dark in mid-winter then very bright with snow as we move towards spring and then depending on the band contribution as vegetation emerges we may then see a band divergence during the late spring into summer. For the southern hemisphere, we have Antarctica, which has more or less an invariant contribution, with sea ice being the only other factor. But with no land areas, we go from bright snow to relatively dark sea for many bands with limited land areas compared to the northern hemisphere changes. The best way of confirming this is to get the full annual cycle for all bands. We should start to see the bands separate out in the late spring and early summer as different behaviours occur in different spectral bands.

From the point of view of tracking changes, the fact that there is a discernible trend at the weekly level strongly suggests we may be able to use weekly analyses for comparison purposes, providing a good deal of detail on any change in the responsivity from one year to the next.

There are discontinuities in the plot in figure 2. Some are major discontinuities in the overall trend and may be due to unusual conditions (such as extensive cloud cover). Environmental effects on the overall averages could be identified by using multiple sensors and/or multiple platforms, so that they can be identified and reduced or removed.

Validation

No validation has been possible at this time, as we need a difference between one year and the next to do some direct comparisons of the weekly trend, so this will have to wait until November 2020. At that time both the calibrator drift and absolute calibration drift can be evaluated in the manner stated in the PVP, in which we can compare direct counts or ratios of counts and calibration data for data collected in the same month of two consecutive years.

Future work

Continuation of the processing of analysis for at least the next six months will be required to validate the effectiveness of the method, based on the validation methods discussed in the PVP. Ideally the work should continue for several years to improve the reference curves and determine how closely the responsivity changes observed for Level 0 match the changes determined from using the on-board calibrator.

Additionally, by using a whole year of data, differences between individual bands should become apparent and each band can be mapped independently.

3. Results - Level 1 OLCI SNR Estimation

The algorithm uses all input Level 1 images and tries to determine both the signal level and noise level for each image based on a moving window method. Accumulated statistics define this signal and noise level for the whole image. Given the heterogeneous nature of the images we find that many of the analyses produce sub-optimal estimates of the SNR. Hence if we plot signal against noise we get a cloud of data points (figure 3, orange data cloud).

The SNR estimate derived from the data cloud is based on the modelled data point that gives the highest SNR assuming a shot-noise limited model. Hence it will be found on the upper edge of the data cloud. In figure 3, there are two curves which use two different samples from the data cloud, based on weekly estimates extracted from the five months of data collected.

The two curves are similar, but clearly the grey curve is capturing a few data points that define the upper cloud edge. The corresponding SNR estimates are not very different at the target radiance of 63 W. An example of the observed differences in SNR using weekly data for November 2019 is summarised in Table 1 and shown graphically in figure 4 for November 2019.

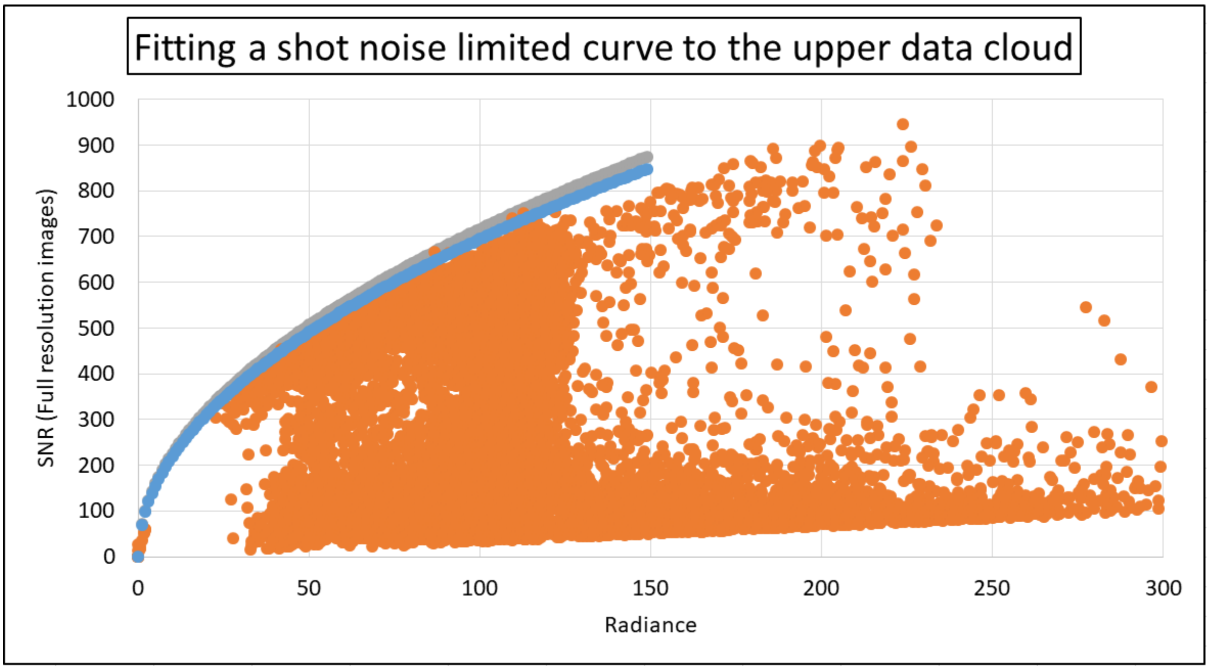

Figure 3: Fitted curves for two different samples from the data cloud

Band	Wavelength	Percentage Difference
1	400	5.76
2	412	3.88
3	442	4.14
4	490	6.32
5	510	1.69
6	560	7.03
7	620	4.01
8	665	3.30
9	674	1.42
10	681	3.47
11	709	5.62
12	754	6.64
13	761	0.33
14	764	0.10
15	768	9.72
16	779	5.05
17	865	7.24
18	885	7.52
19	900	1.63
20	940	4.84
21	1020	2.57

Table 1: Maximum percentage difference between four sets of observations in figure 4 (These are max/min values, percentages >5% are highlighted in orange)

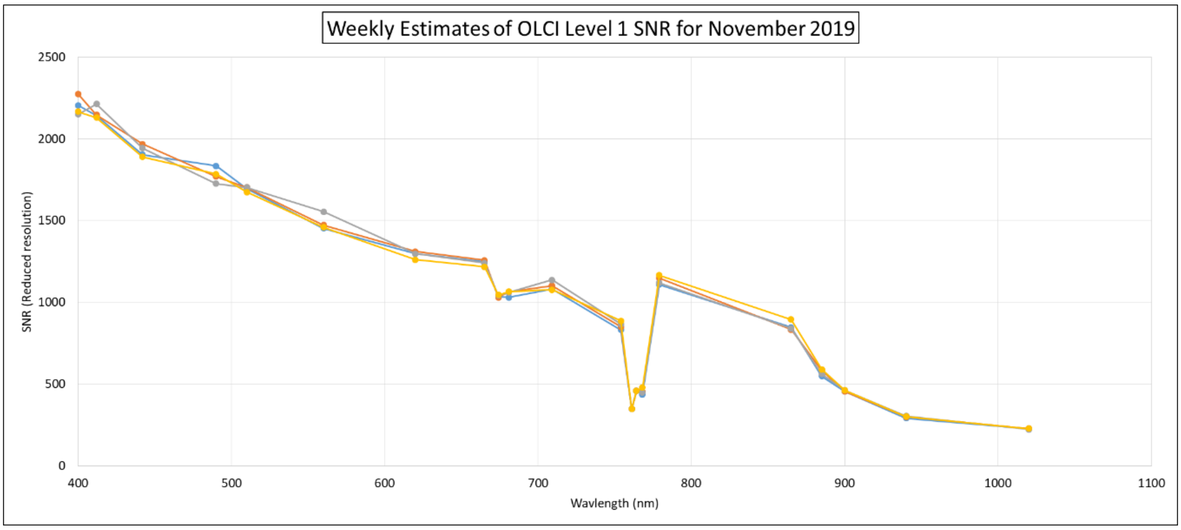

Figure 4: Weekly estimates of OLCI level 1 SNR for Camera 1 from November 2019.

Overall it seems, we get very good and consistent estimates of the SNR using weekly analysis of the data provided by Eumetsat as shown in figure 4. Even daily data (figure 5) can produce reasonably consistent data. The output values used in figure 4 and figure 5, were generated using the analysis algorithm for SNR for OLCI provided to Eumetsat. For any given spreadsheet from the SNR databases from OLCI we can derive the SNR for all 21 bands at the specific target radiances for each spectral band.

An example showing the results for single days of different months in 2018 are shown in figure 5.

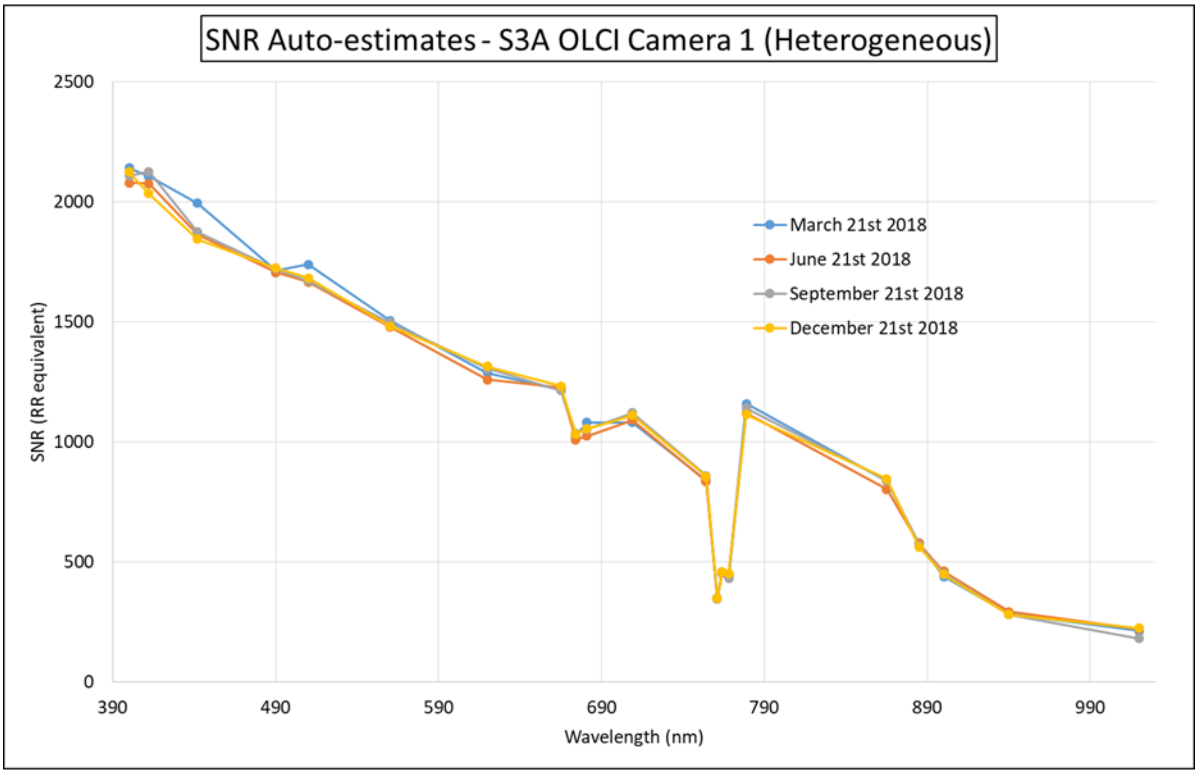

Figure 5: Daily estimates of SNR from four different days in 2018.

As can be seen in figure 5, the results are very consistent from one month to the next, they are even highly consistent from one week to the next. However, they are different from those derived from the diffuser (figure 6).

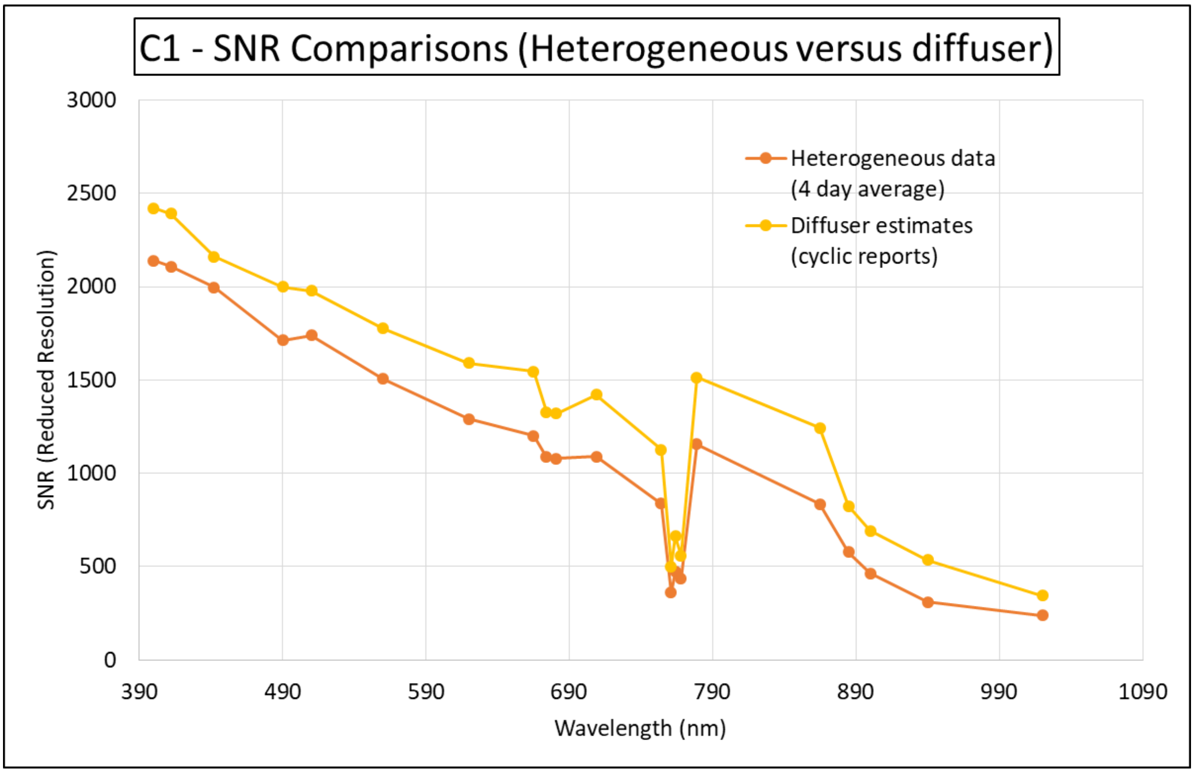

Figure 6: Comparison of diffuser data (yellow) and our estimates of SNR

This disparity will be discussed in detail in the validation section for this algorithm.

Another feature of the repeated measurements of this technique is that we can see temporal changes which do not seem to be observed by the on-board diffuser (figure 7). In figure 7, where the x-axis is now band number not wavelength as seen in previous plots, we have compared data from approximately one year apart - from December 2018 to November 2019. Although much of the SNR plot is almost identical, there is a deviation between the older data which has lower SNR values and the more recent data which has higher values in the first four spectral bands.

As we have discussed when describing the algorithm we use, it is based on small spatial windows (up to 7x7 pixels), which means there is a contribution to the effective noise from any differences in the relative gain between neighbouring detectors.

We noted that in April 2019, between the two sets of observations shown in figure 7, there was an update of the relative gain model, which reduced the observed relative gain variations in the shorter wavelength bands significantly. This change is discussed more in the relative gain section. This change has had a distinct impact on the SNR values for these shorter wavelength bands.

One question we need to keep in mind, is why only these specific bands were affected by the change.

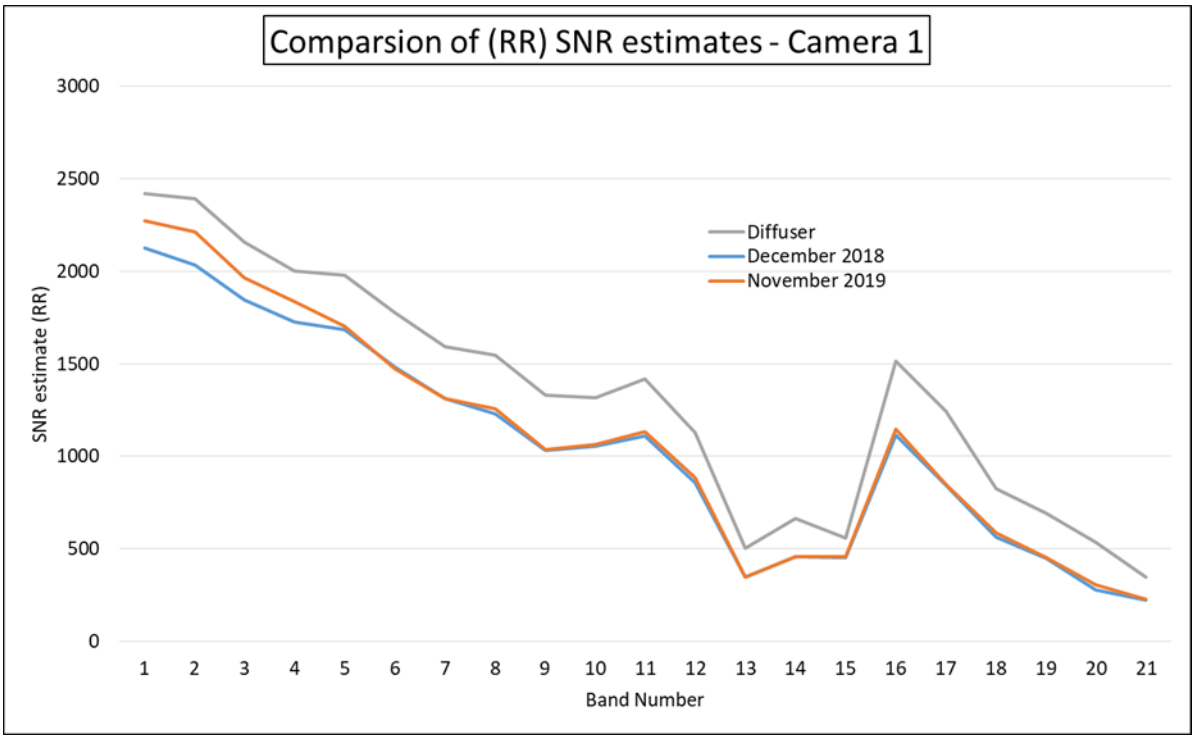

Figure 7: Different representation of the SNR data by band rather than wavelength

In figure 7, we can see a large change in the first four bands of OLCI in terms of SNR between December 2018 and November 2019. During that period the relative gain model was modified and small residuals seen in the imagery were reduced, especially for Band 1. In figure 8 on the following page we show a simple diagram consisting of four plots.

The x-axis of these plots is the detector number and the y-axis is the magnitude of the relative gain residual. The scaling for all four plots is the same. If we look at the residuals, it is clear that as we move from March to June, to September and finally December 2018, that the residuals are growing in magnitude.

It is also clear that certain relative gain detector value differences are very strong and persistent from one plot to the next, especially from June onwards. These values had been increasing for some time before 2018. During the April 2019 relative gain model change, the values were strongly reduced, producing a corresponding improvement in the effective SNR.

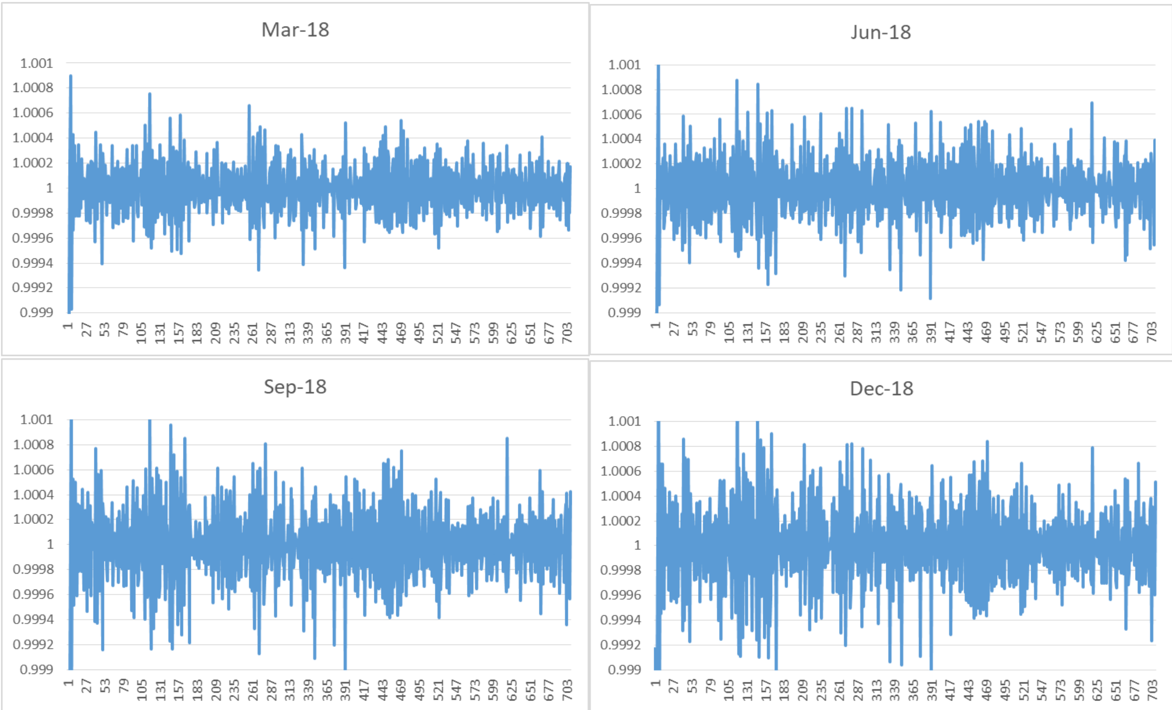

Figure 8: Observed residuals in Camera 1, Band 1 for four separate dates of data collection

The SNR determined using our windowed method will depend to some degree on these detector to detector differences. In April 2019 when the relative gain model was modified, most of these features in figure 8, were strongly reduced in magnitude and this produced the drop in SNR in figure 7.

So the SNR is providing a lot of information on the system behaviour. As the number of data points is increased, for example using all five months of data from November 2019 to end of March 2020, the SNR clouds become highly populated and we can now clearly see the upper edge of the plot defining the SNR cloud (figure 9) and see differences between different Cameras for the same spectral band of OLCI.

In this case Camera 1 is showing higher SNR than Camera 5. The red star is the diffuser estimate at the target radiance for Camera 1. Our data suggests that the diffuser tends to give a higher SNR than that from other sources (see the validation section for more details).

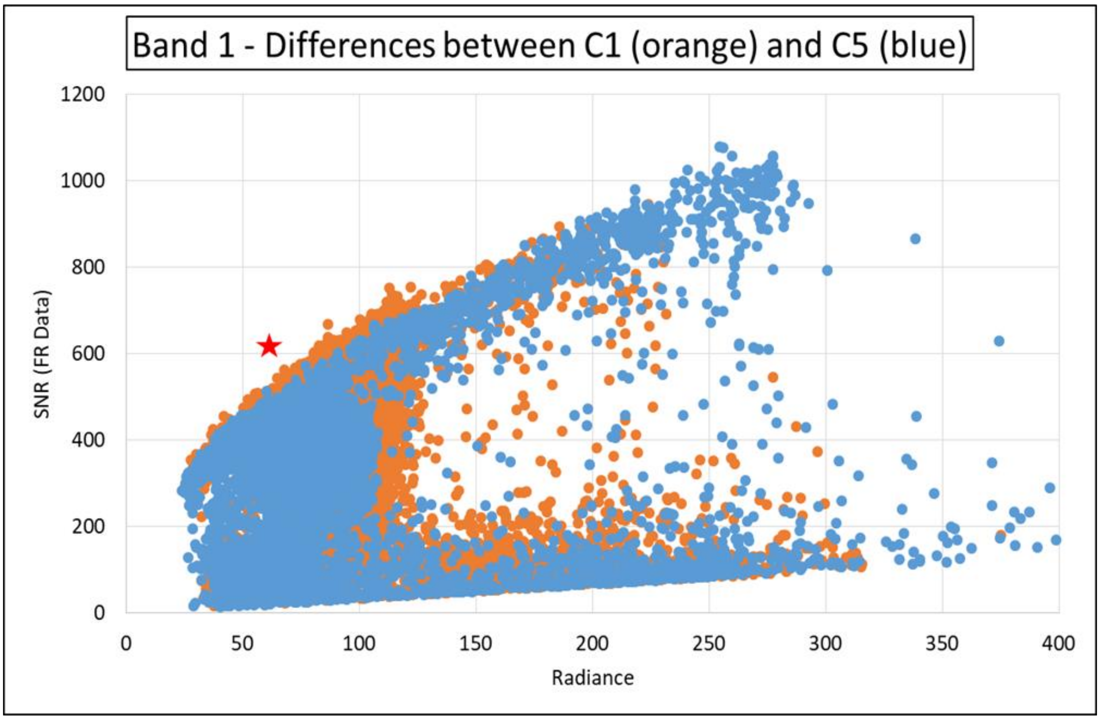

Figure 9: Comparison of band 1 SNR for cameras 1 and 5.

One issue of using this methodology relates to the quantity of data used. In figure 10, we show SNR data clouds from band 21 for the four single days in 2018 (different colours) and the SNR curves from different model fits. The impression is that the data is not shot noise limited at the target radiance (blue vertical line) and is well below the diffuser estimate, value of 86 on the y-axis.

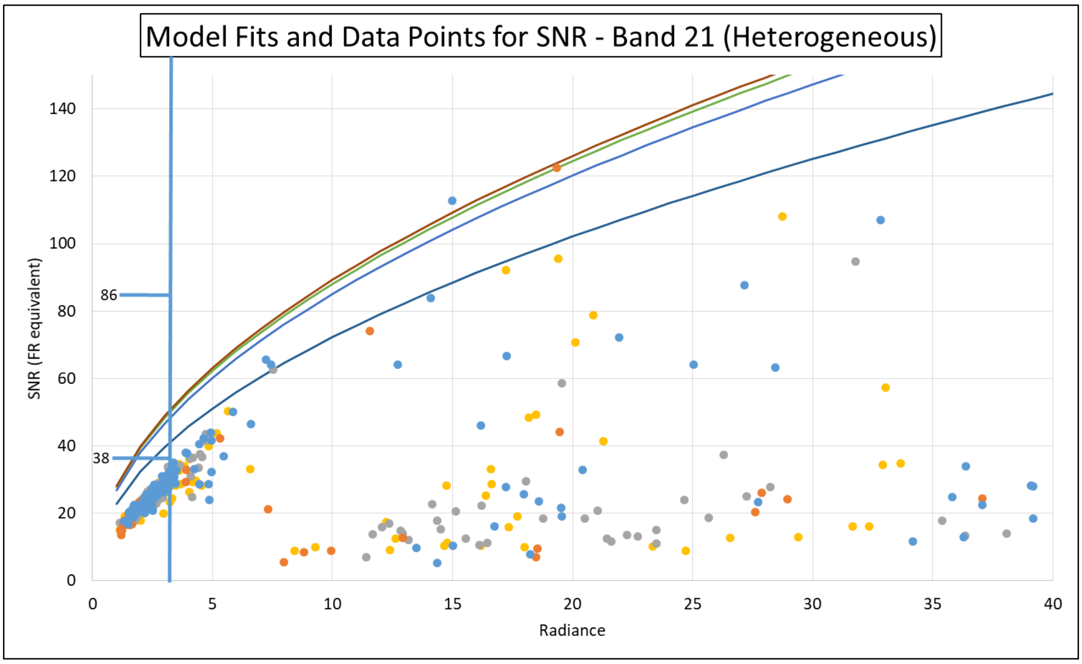

Figure 10: Data cloud and fitted curves for band 21 compared to diffuser and requirements.

However, if we add enough data, the expected shape of the data cloud emerges (figure 11) and we can model the SNR more effectively. The target radiance is the red vertical line in figure 11. Note that the diffuser value (red horizontal line) is still very much above the upper boundary of the data cloud, suggesting that the SNR estimate from the diffuser is a poor estimate of the reality when considering spatial averaging at these very low radiances. Again, we can track this back to uncorrected relative gain differences as a potential source of some of the observed differences, as discussed in the validation part of this section and the relative gain discussions later in this document.

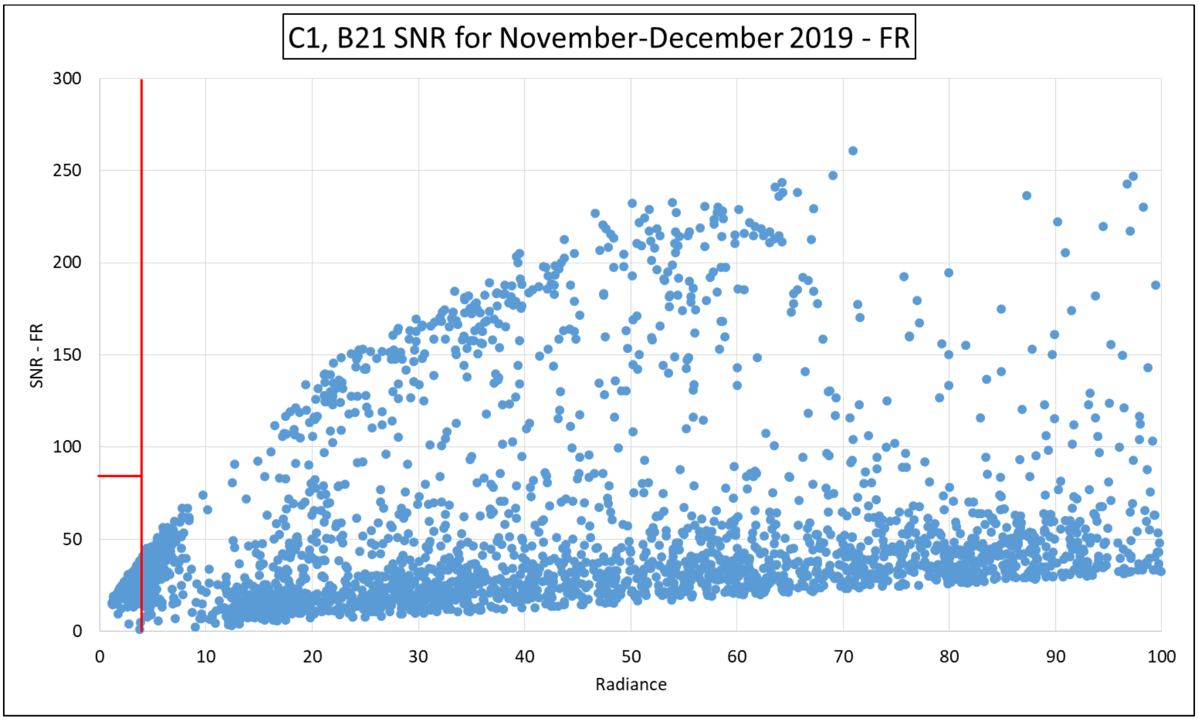

Figure 11: Camera 1, band 21, SNR cloud showing shot noise behaviour.

Validation

The primary means of validation stated at the start of the project was to use the on-board diffuser. We expected a slightly lower SNR than the diffuser estimate for SNR as we assumed the SNR was calculated on-board at the detector level rather than spatially, although we do not have any confirmation on how exactly the SNR is calculated on-board. This could be key information required to determine why there is a discrepancy between the diffuser based and our method for evaluating SNR.

However, results showed that there was a large difference up to several tens of percent between the diffuser estimates of SNR for Reduced Resolution data at the target radiances (from cyclic reports) compared to our estimates using heterogeneous images (figure 12). Even after the relative gain model was changed, only the first four spectral bands showed a marked improvement in SNR, but still not enough to bridge the gaps to the diffuser measurements.

This raised the question of what was the cause of the disparity. Certainly a large part of it is related to our use of spatial data covering several detectors, compared to what we assume to be single detector estimates made using diffuser data. However, as shown after the relative gain model was changed, essentially eliminating the relative gain effect for the first four bands, we found that this change accounted for only half of the difference in SNR observed in figure 12, suggesting that there is another issue.

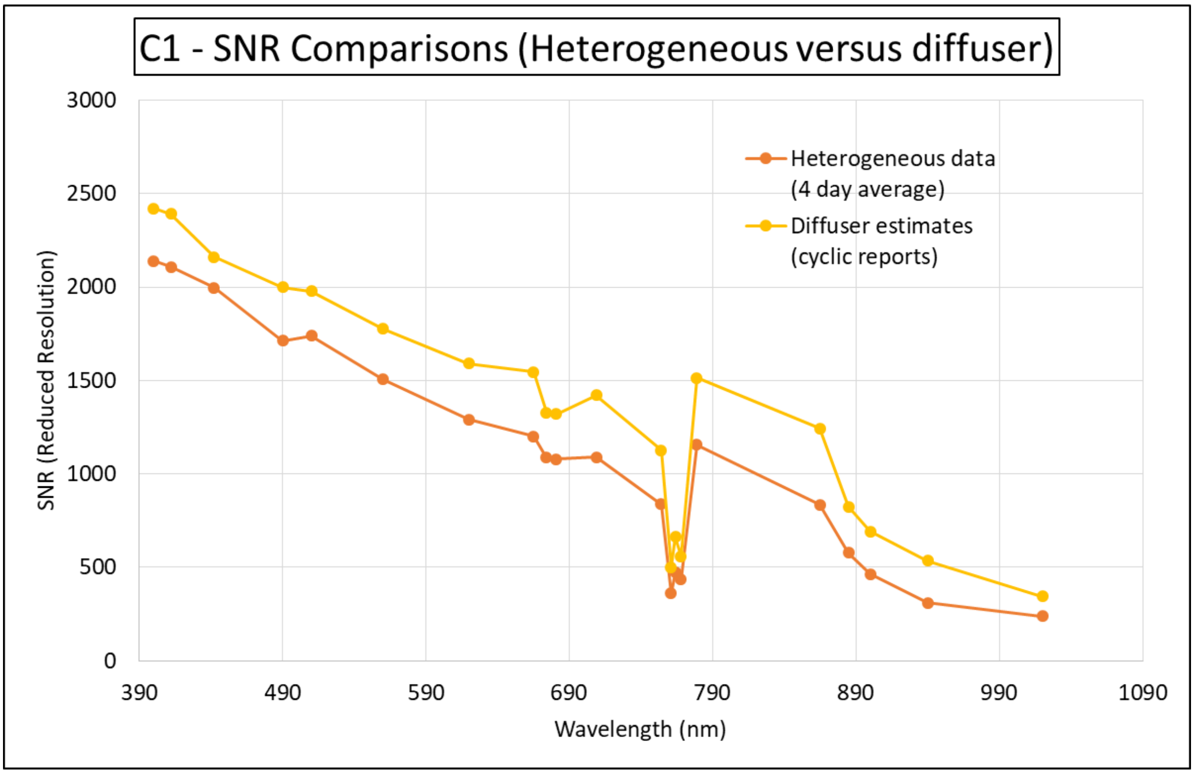

Figure 12: Heterogeneous data compared to diffuser estimates for SNR for OLCI C1.

The only recourse we had was to look for other information and methods that might establish which is the more likely true estimate of the SNR. The only two possibilities we considered was to use pre-launch data to see how well that compared to our results and to use snow scenes. Given their brightness and almost homogeneous surface, we would expect a very good estimate of the SNR.

Pre-launch data proved very difficult to find. Only a single plot and a brief description were available in one document [AD2]. Data points were estimated from the plot and are shown alongside our heterogeneous estimates in figure 13.

The other alternative was to use snow scenes. We therefore collected 14 snow scenes (which it should be noted is a very small number of images). These images were cloud free and showed good homogeneity. The SNR was derived based on a modelling approach developed by Dr. Settle, rather than using the same windowed method we used with the heterogeneous images. These results are also shown in figure 13. Even though the number of images is small, the one characteristic of homogeneous images is that a single good image can give a very good estimate of the SNR.

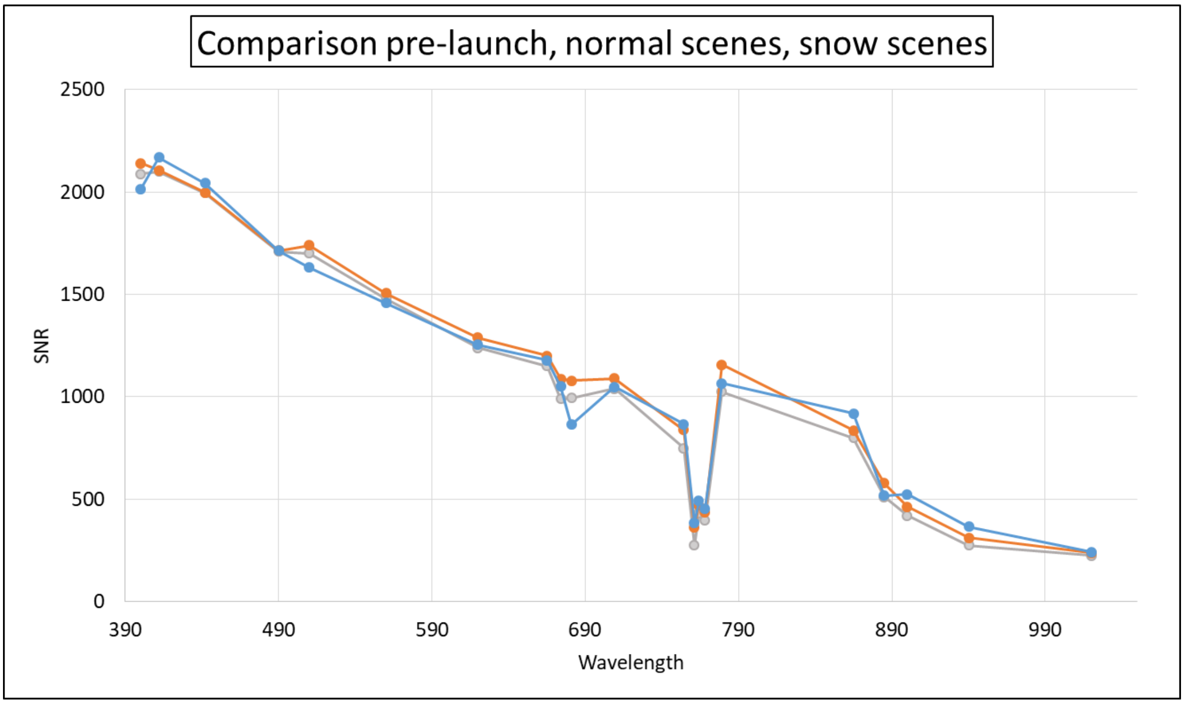

Figure 13: Comparison of estimates of SNR from pre-launch (grey), heterogeneous (orange) and snow images (blue).

The results in figure 13 show that both the homogenous scenes (using spatial statistics) and the pre-launch data (methodology not known) both give SNR values more in line with the heterogeneous method than with the diffuser estimates.

There are many things we do not know in this study:

- How exactly are the SNR values calculated for the on-board diffuser? What are the steps? Is there any modelling that requires smoothing?
- What were the exact values obtained pre-launch for the instrument? The values must exist, pre-launch testing must have taken place. How were the measurements made?

With this basic information, we could probably determine why we see differences between these different measurement sets. We can only assume at this time that the heterogeneous, snow and pre-launch are based on spatial measures and that the on-board measurements are not and hence we can see some difference in the estimates due to this, although this would not explain the total difference observed.

Future Work

The major issues that still need addressing are in the area of validation. The rather large differences between the diffuser measurement and the other methods employed needs explanation. Therefore the most important first step is understanding how the measurement is taken with the diffuser. What models are used? How is the value actually determined step by step? The next area is the pre-launch, it seems surprising that no adequate pre-launch assessment of the SNR took place. However, the approach to recover this information would best be made from Eumetsat to ESA.

Once these issues have been resolved it will become clearer on what is the true effective SNR of the sensor. However, it is clear that the methodologies perform well, subtle variations between cameras are present in the diffuser and heterogeneous data and the overall band to band relationships are clearly seen. This discussion will be continued in the summary and discussion section later in the report.

One other item worth pursuing in the longer term is that of assessing any gradual drift in the SNR values. Repeated observations, I would suggest a rolling month average, would show any gradual drift in the values for any band and this can then be related directly to any other issues which might affect the spatial averaging such as relative gain changes with time.

4. Results - Level 1 OLCI Relative Gain Estimation

We identified in the Product Validation Plan that we can consider two types of relative gain residual,

- 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 for the diffuser on-board they are not visible. These can persist for the lifetime of the sensor if not detected and removed.

In theory the algorithm can derive the relative gain residuals from a single image, which is very useful for observing the emergence of anomalous changes in the relative gain of individual detectors (Type 1 above), where a large change produces visible striping in the imagery. However, the best approach at extracting the very small persistent residuals, the type 2 residuals above (smaller than 0.05% of the signal level in some cases) is best done using an average residual derived from multiple images.

We have had a sufficiently large dataset to evaluate the relative gain of the OLCI sensor. The diagnostic data sets from four different dates in 2018 showed that the features derived from different dates are very consistent (Figure 14). This is a subset of detectors (x-axis) but is typical of the types of features present and their persistence with time.

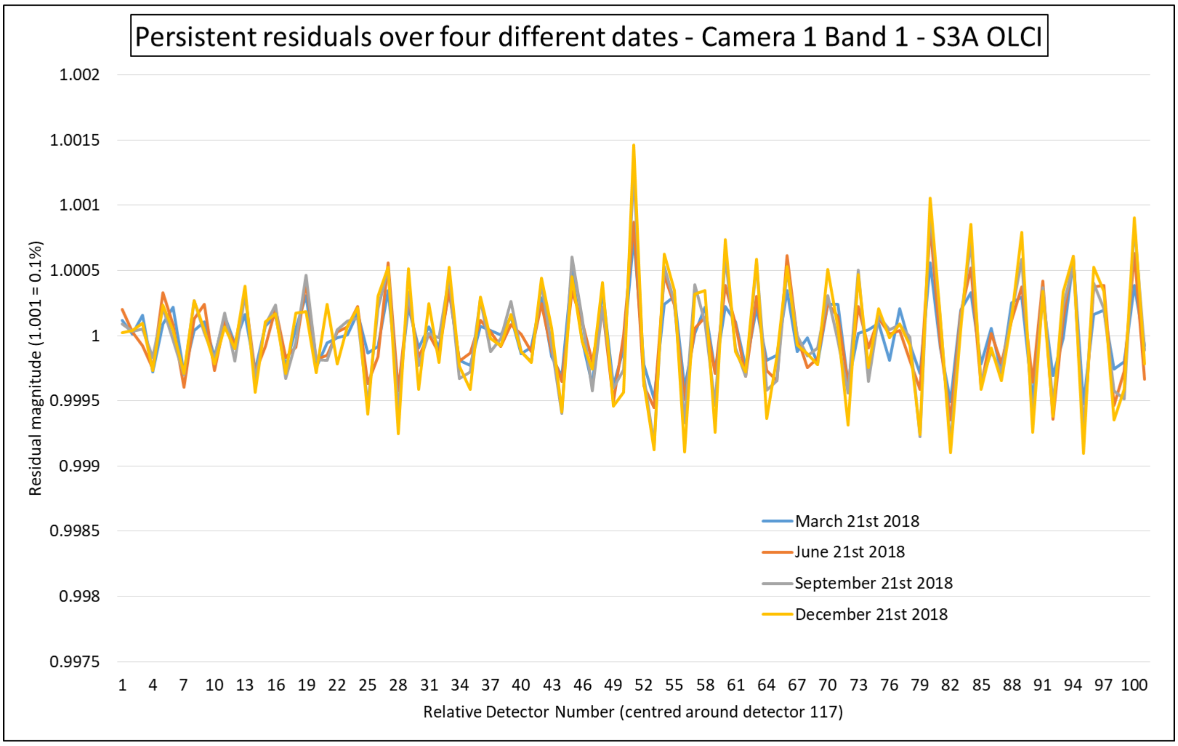

Figure 14: Residuals around detector 117 showing consistency with time

But also as shown previously in figure 8 in the SNR section, the magnitude of these persistent residuals was increasing with time in 2018. In fact looking at figure 14 we can see that the orange line is nested inside the yellow and grey lines typically. So June residuals are less than September, are less than December.

In April 2019 the relative gain model was changed - the point of change can be shown by mapping a single detector response with time (figure 15). The residual magnitudes for some bands, for example band 1, decreased significantly with this change (figure 16).

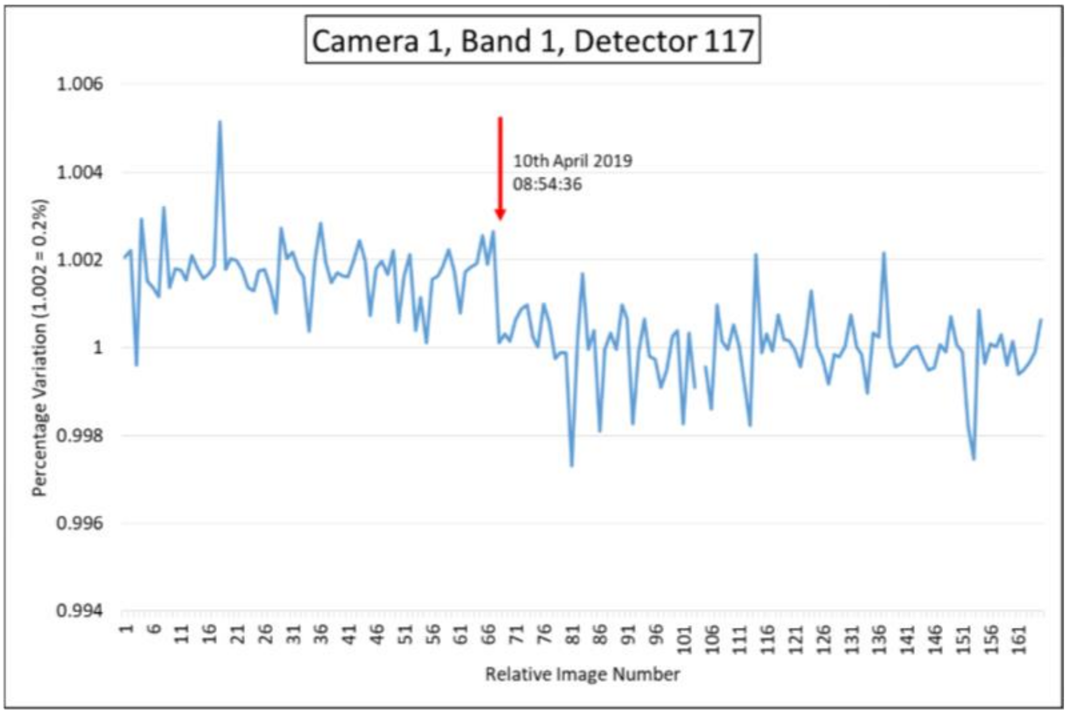

Figure 15: Change in a single detector response due to relative gain model changes

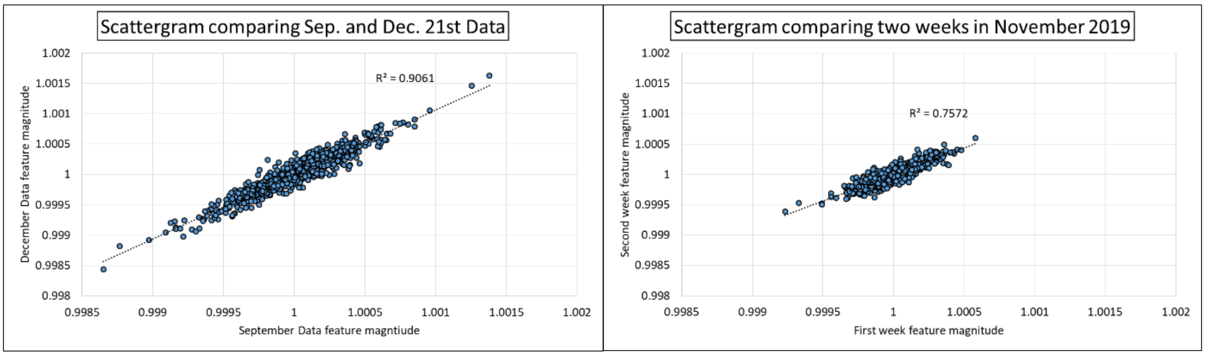

Figure 16. Scattergrams from observations for band 1 before and after the relative gain change

In figure 15, the Relative Gain Model was changed according to the cyclic reports on April 11th. We actually saw the effect of the change on the 10th April at 08:54:36 UTC, with a specific detector that had a feature running at around 0.2% (1.002) of the signal level, so just about visible as a stripe, resets to essentially no feature (a value of 1.0).

In figure 16 we can demonstrate the effect it had when comparing two data sets taken before the change (left) and two data sets taken after (right). The scaling on both axes is the same. The x and y axes are the magnitude of the persistent residuals observed. As can be seen the magnitude is more than halved, the correlation in this case is also weaker, but still quite strong. This suggests that although the relative gain model change reduced the magnitude of the features for band 1, the features that were left were still persistent and repeatable from one image to the next.

It was noted that the relative gain values for some bands remained unchanged (figure 17) or even increased in value after the relative gain model was changed (figure 18).

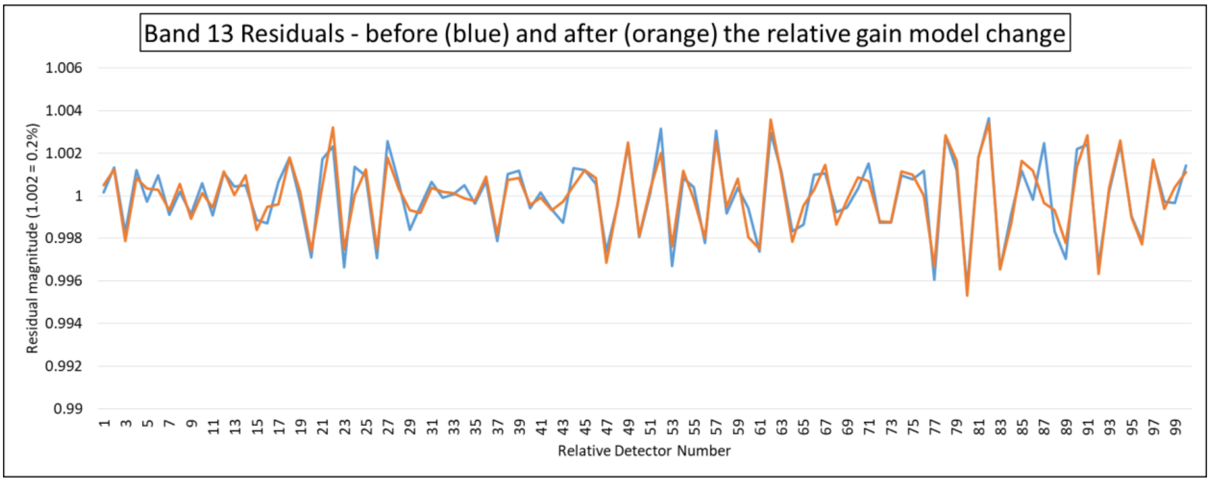

Figure 17: Comparison of relative gain values before and after the relative gain model change

The figure 17 comparison is the before and after comparison for Band 13. The features are essentially identical with no change in magnitude.

In figure 18 we see an even more unusual set of behaviours. The left plot in figure 18 is comparing the September and December 2018 data in blue and grey lines.

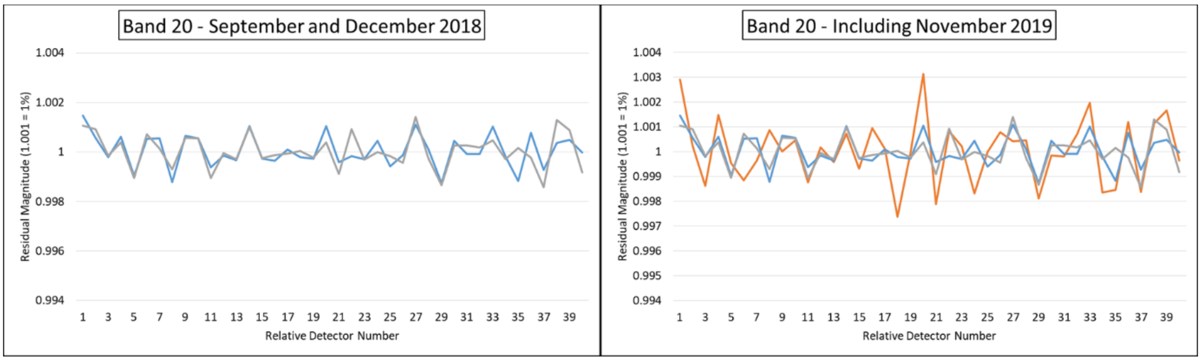

Figure 18: Left plot shows pre-model change results, right plot overlays the post-model results.

The right plot has the November 2019 data added to it. The features are more or less in the same place, but the magnitudes in some cases are much larger. We considered that this magnitude change might be due to different brightness targets contributing, as the depth of the features may be due to non-linearity effects, as discussed in the next major section. However, when we compared all the data we have after the model change, we found that the weekly results over a five month period were very consistently higher than before the model change.

In other words, this looks like a real change in the values producing more striping in the Band 20 output after the new relative gain model was employed. An unexpected result.

The reason why some bands show significant changes in the persistent residual magnitude after changing the relative gain model and other bands show no real change may be due in part to the origin of the persistent residuals themselves. The assumption discussed throughout this document, is that since we can see these residuals in normal images at normal earth radiance and cannot see them with the on-board diffuser that they may be related to non-linear effects. This will be discussed further in the non-linearity section which follows.

However, although we considered initially that the relative gain correction was a multiplicative correction, we saw features in some bands that were a mixture of multiplicative and additive in some cases, in other bands almost pure additive. If the correction is only multiplicative then only some of the bands will improve. Those with a large additive term or almost purely additive term would not be reduced and the features would remain equally strong or as shown in the example for Band 20 increase, after a relative gain correction.

One of the analyses we performed towards the end of this project was to reduce the data down into a manageable form to identify those detectors that were causing the most issues. This was carried out rather simply, by ranking the magnitude of the observed residuals and determining the cut-off boundary points which encompass 90% of the detectors. In figure 19 we show an example plot.

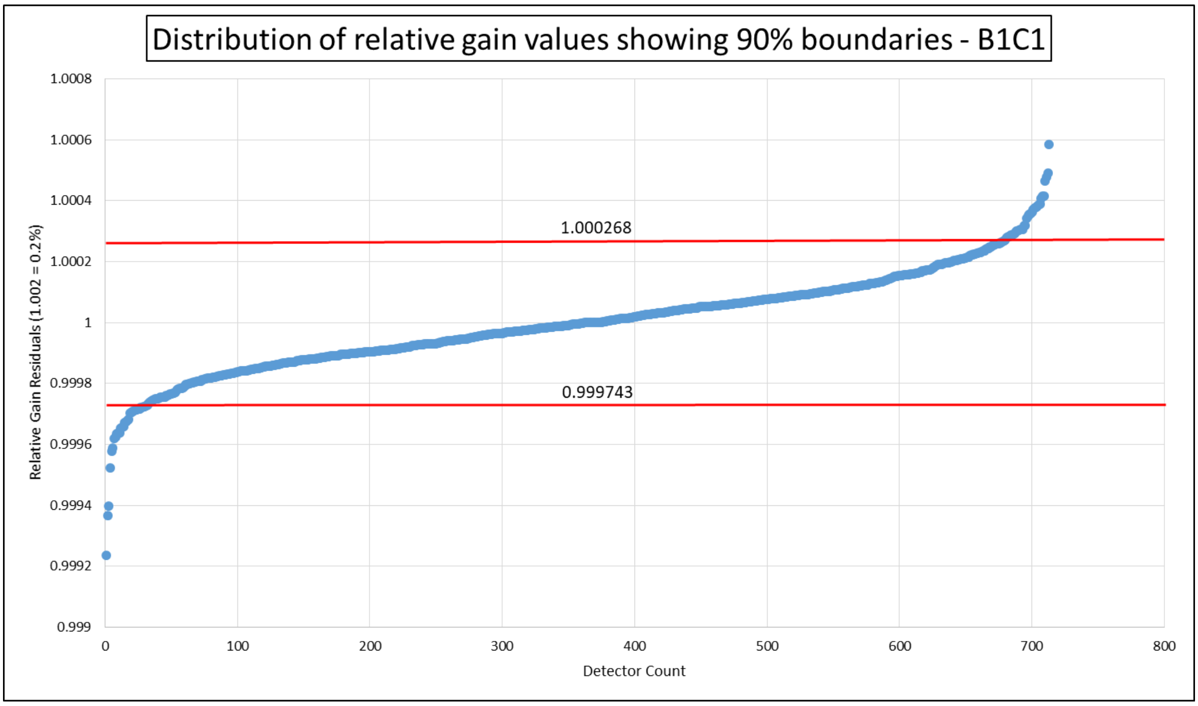

Figure 19: Band 1 Camera 1, relative gain residual magnitudes with 90% of data between red lines

In figure 19 we can see that 90% of the residuals have a magnitude of around 0.026% or less, so very small. We have carried out this analysis for all bands and the results are shown in table 2.

In table 2, the lower bound and upper bound values are those defined by the red lines as shown in figure 19. The last column, the Average Absolute Residual, is essentially the average residual across all detectors in absolute terms. The bounds limits and absolute residual give similar but slightly different results, as the average absolute value is dependent on the shape of the profile as shown in figure 19. A flat line with sharp drop and rise at each end would have a low average absolute value, while an inclined profile will have a higher average absolute value.

Band	Lower Bound (90%)	Higher Bound (90%)	Average Absolute Residual
1	0.999743	1.000268	0.015%
2	0.999749	1.000276	0.016%
3	0.999658	1.000346	0.02%
4	0.999647	1.00039	0.023%
5	0.999418	1.000574	0.031%
6	0.999628	1.000365	0.026%
7	0.999512	1.000442	0.033%
8	0.999642	1.000313	0.032%
9	0.999608	1.000362	0.033%
10	0.99959	1.000331	0.035%
11	0.999626	1.000295	0.034%
12	0.999573	1.00035	0.038%
13	0.997622	1.002171	0.113%
14	0.998124	1.001886	0.095%
15	0.99917	1.000727	0.052%
16	0.999578	1.000309	0.041%
17	0.99935	1.000512	0.055%
18	0.99917	1.000565	0.060%
19	0.998929	1.000794	0.068%
20	0.998067	1.002074	0.113%
21	0.996639	1.0036	0.183%

Table 2: Variation in the residual magnitude by band for Camera 1.

The results in table 2 for the bounds are shown graphically in figure 20.

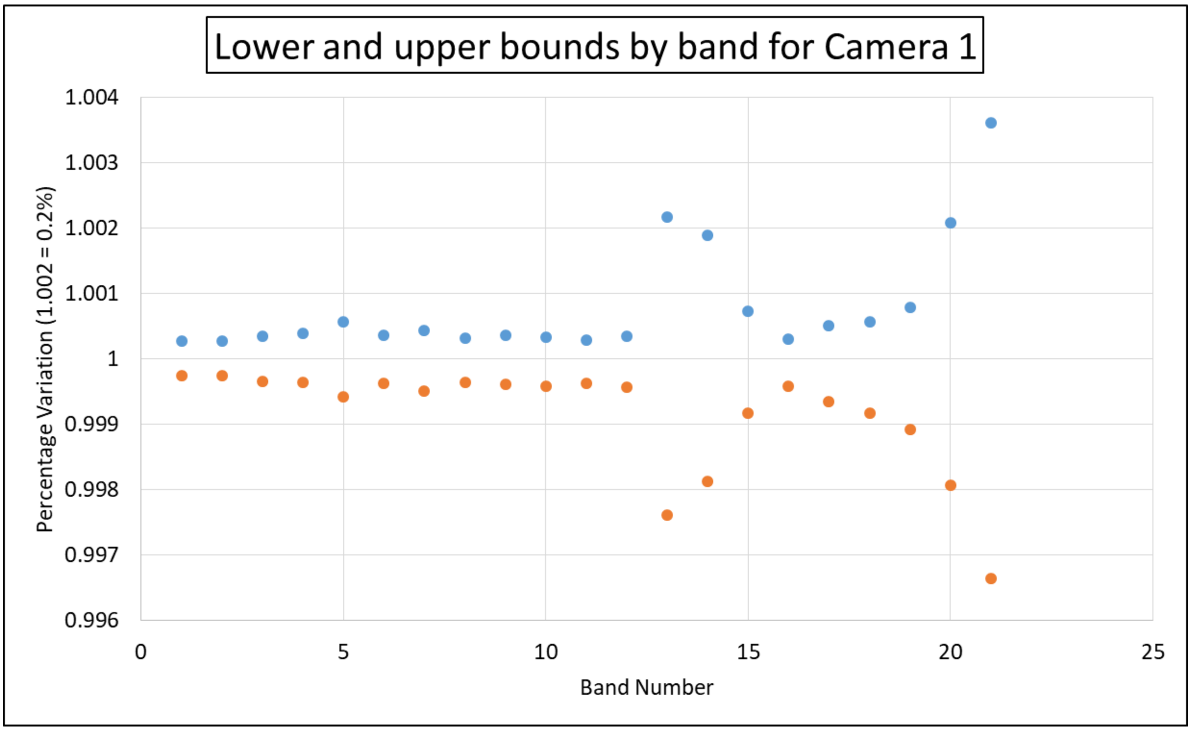

Figure 20: Bounds in which 90% of the data falls, all bands, camera 1.

In figure 20 on the previous page the bounds increase with band number up to band 5, then drop slightly and stay more or less stable until band 13, where this a very large change. Assuming we are concerned with striping in the imagery, if the residual exceeds 0.2% of the signal level we can start to see striping over homogeneous surfaces. In figure 20, over 10% of the data exceeds that value for bands 13, 14, 20 and 21. For band 21, almost 10% of the data exceeds 0.4%.

In figure 21 below, we have also plotted the average absolute residual. As mentioned, this is more of a measure of which bands show the highest average variation, rather than the most extreme variation.

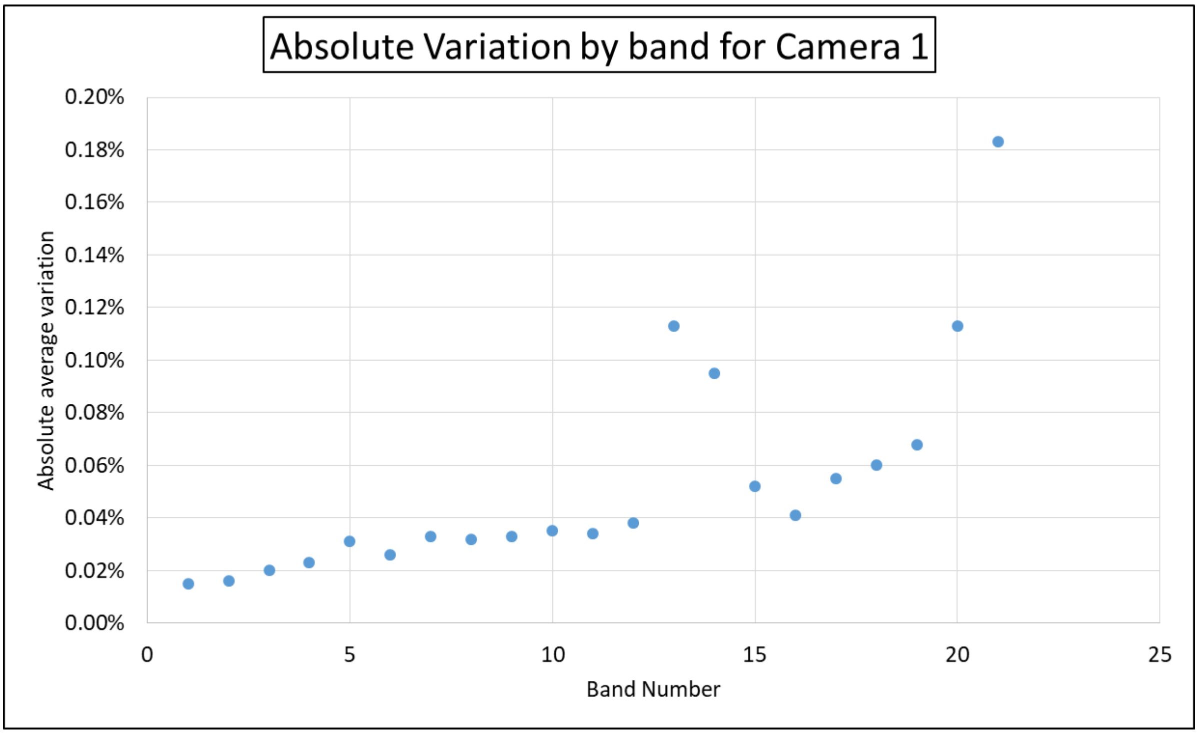

Figure 21: Average absolute variation by band for camera 1.

As can be seen in figure 21, it is similar to the bounds plot of figure 20. However, there are some subtle differences. We can see if we ignore the large spike in band 13 and 14 that the overall trend is increasing persistent residual size as we move from shorter to longer wavelengths. The average value of the residuals itself is small (y-axis is percentage of signal). Again we see a discontinuity at Band 6 - the cause is unknown. However, as Band 5 is 510nm while Band 21 is 1020nm, we considered it might be related to preventing second order contamination of longer wavelength bands. However, until we know more about any placement or use of blocking filters we can only speculate at this time.

The trend on the residuals increases slowly until we hit the narrow atmospheric bands, where there is a dramatic increase in the overall magnitude of the residuals. The values after the narrow atmospheric bands continue to show increasing average residuals, with bands 20 and 21 showing particularly large residuals.

In summary we can see which bands show the largest residuals and also identify the 10% of the detectors showing the strongest residuals. This analysis has helped in identifying individual pixels to examine when we discuss non-linearity in the next major section. In this section we have discussed the presence of persistent residuals. However, we have not yet asked the question why we can see these residuals and yet the MPC cannot see them.

MPC reports point to very low levels of relative gain variation and whereas in some cases we see features exceeding 1% (in Bands 20 and 21), the MPC does not register them. Which raises the very important question of why this is the case.

It is recognised that the MPC on-board measurements for OLCI are generally confined to the very bright diffuser and the dark shutter. If a suitable non-linearity correction is applied, with these two measures the calibration is more or less guaranteed. However, if there is some issue, for example stray-light in the sensor, second order effects from the diffraction grating or some change in behaviour that affects either the bias term subtracted from the data or a gain term. Then we could find that the detector response may vary in the radiance range between the dark shutter target and the bright diffuser target, leading to detector differences at specific brightness levels, leading to striping in the imagery, which is effectively a relative gain issue. These would be our persistent residuals.

This will be discussed in more depth in the non-linearity section which follows, but this would explain why we see features at background earth radiances that are not detected by the on-board sensors.

<u>Validation</u>

The original idea we had was to use the diffuser to identify the same features we could identify (the Type 1 features). However with Type 2 persistent residuals, they will always be present as the diffuser cannot be used to identify them and hence remove them.

There is in fact no suitable alternative than to examine the imagery alone. Hence the only proof we can provide is the visual presence of striping at locations we can identify by examining the top 10% of very strong residuals we see in the relative gain outputs. However, we have to be careful that we look in the right radiance range. Some radiance values may show no residual, due to variable response with surface brightness. The ideal would be to have a set of homogeneous targets at different radiances, but this is not really feasible.

Another alternative that can be considered is to collect side-slither images. Then within a single image you will sample a range of different radiances. Neighbouring detectors can be compared and a corresponding set of comparative changes extracted. This may not give full non-linearity curves, but should give a good approximation of how the signals are changing relative to neighbouring pixels.

Since we do not have a full set of data for validation, we instead processed a small subset of images to show the presence of striping induced. The initial results of this validation proved very promising. If we took the strongest features identified in the relative gain results for a specific band and camera we could always find examples of their presence in the imagery.

A very good example as expected is over a relatively homogeneous water body (figure 22). This is an extract from band 21 of image,

S3A_OL_1_EFR____20200101T135515_20200101T135815_20200101T154636_0179_053_181_3600 _MAR_O_NR_002.SEN3.

This is one of the images provided by Eumetsat for our visual validation.

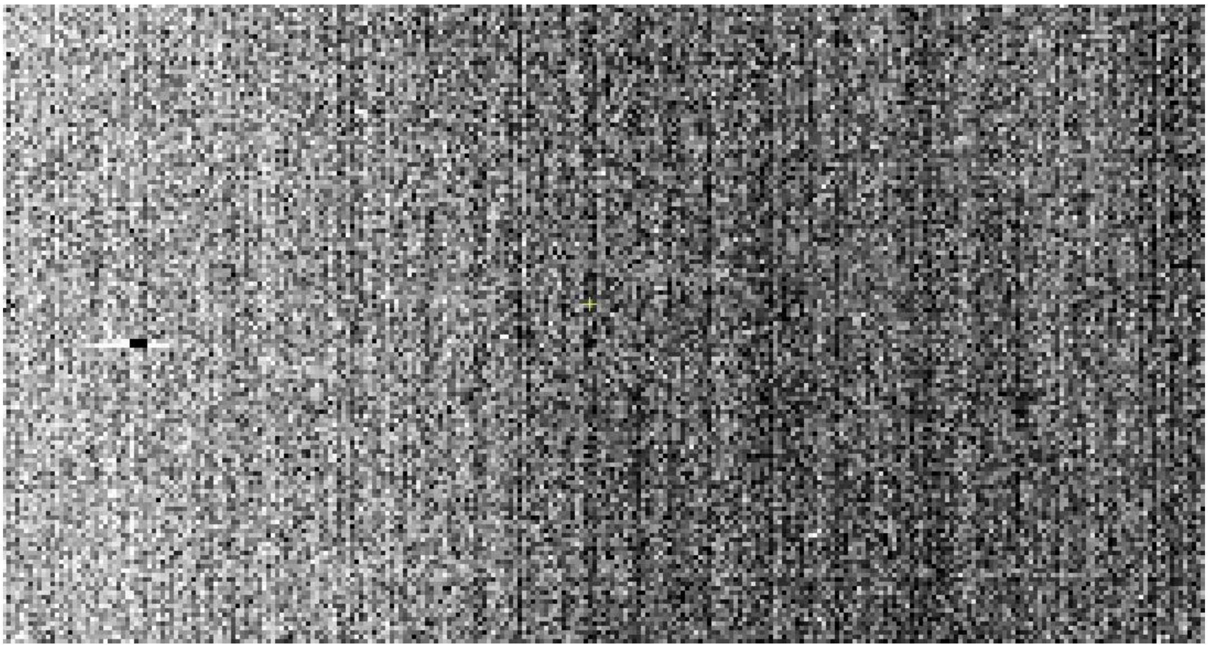

Figure 22: Subset of an image over water for Band 21, Camera 1

Each of the major dark and bright stripes seen can be found in the relative gain output. The stripes extend for as long as there is clear water but fade (as expected) as they enter very cloudy and coastal regions, as they are mainly confined to low radiance targets.

Band 20 shows similar behaviour to Band 21. We tracked each feature as far as possible until they faded into areas of cloud and land. One thing we did note is that they faded far more rapidly than we anticipated based on non-linearity plots (see next section on Non-linearity). However, this can now be explained and details will be given in the non-linearity section.

The second highest residual for Band 21 is at detector 602 of Camera 1 (figure 23).

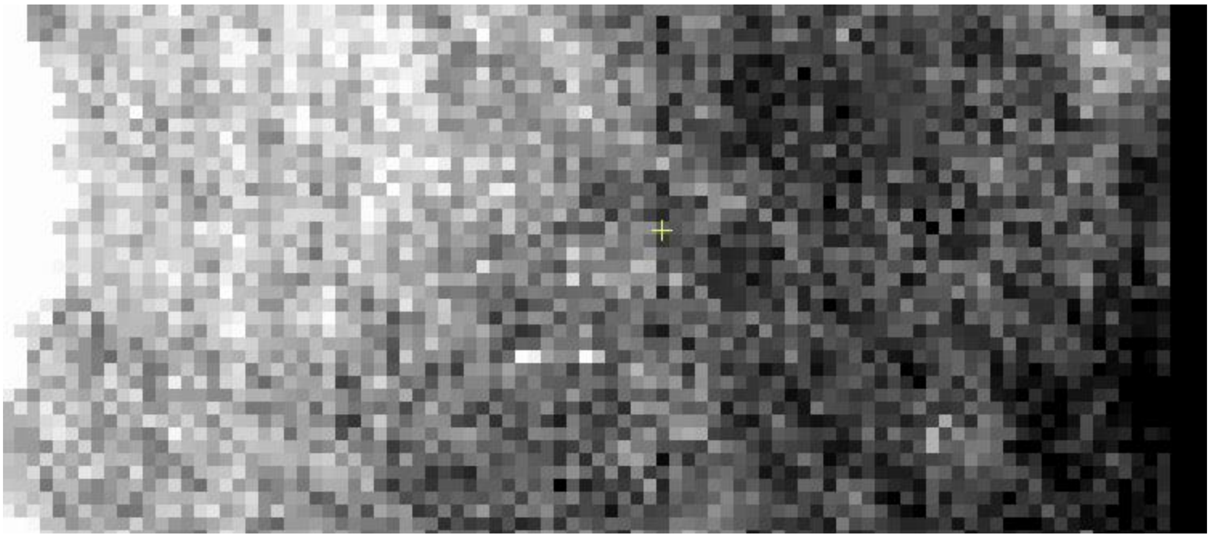

Figure 23: Dark line under the cross-hairs shows the detector producing a large residual in Band 21.

For numbering purposes, residual 1 is the residual of ratioing values from detector 1 and detector 2. So residual 602 is the ratio of values from 602 to 603. The detector identified in the imagery is detector

603. So we are ratioing a bright set of values from detector 602 to those of a dark set from detector 603 giving average residual values of 1.008 or 0.8%. This of course is only part of the picture as the expected residual values should vary according to radiance if non-linearity is present.

As a very simple test of the difference between detectors 602 and 603, we looked at the column values in a small area (17 along track pixels) for each column. The difference in signal level (at around 4W) was 1.8% of the signal level, which is higher than the average value from our non-linearity estimates. Again this apparent disparity can be explained. Consider that the average residual is over all brightness targets and from our knowledge of how the residual varies with image brightness, the residuals are larger at lower radiances, explaining the difference observed.

Similar stripes were identified for all the detectors showing large residuals for Bands 21 and 20.

For band 1, only the largest residuals could be identified as shown in figure 24. This is at Detector 667 and shows an anomaly at 0.08% of the signal level based on the data from the relative gain analysis. The features in Band 1 were very difficult to observe, but given we normally use a rule of thumb for visible features of 0.2% it was surprising to see this feature, especially given the background radiance was quite high at just under 100W.

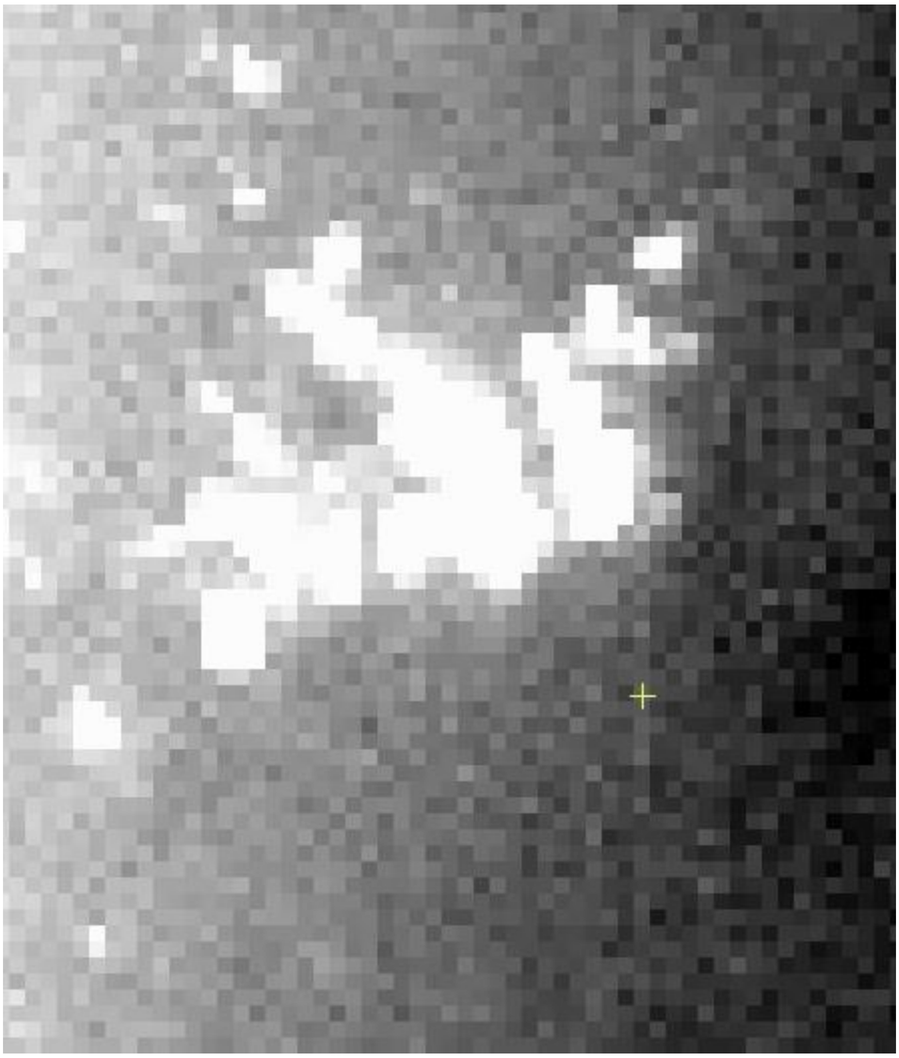

Figure 24: Weak bright line in band 1, camera 1.

For band 13, we looked at the top 10 dark stripes and bright stripes (Table 3). Note that detector 633 is the darkest stripe and 634 is the second brightest stripe, so neighbouring pixels.

Dark Stripes		Bright Stripes	
1.003025	361	0.995411	83
1.00306	355	0.995801	634
1.003079	65	0.995867	337
1.003109	338	0.996442	23
1.003221	490	0.996502	26
1.00327	25	0.996542	214
1.003318	458	0.996602	501
1.003374	373	0.996689	171
1.003485	85	0.996763	95
1.003511	633	0.996792	86

Table 3: Darkest and brightest stripes in the imagery

The ratio between these two columns (633 and 634) would be higher than normal given the brightness differences. So it is important to be aware of that both columns contribute to the overall magnitude of the observed residuals.

Band 13 proved a very useful band to work with, as it is possible to see variations in relative gain across a large radiance range with magnitudes that should be visible. In figure 25 we see an example over cloud with radiance levels at 186 watts. The feature is barely visible under the cross-hairs as a slightly darker line with a signal variation of 0.4%.

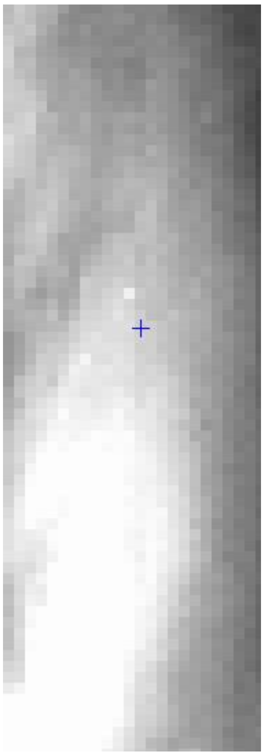

Figure 25: Band 13 feature at detector 85 over cloud.

Some Band 13 features were quite difficult to detect - consider that the features we are looking for are typically around 0.4% of the signal level based on the non-linearity plots. We found we could track single detectors for very large distances over homogeneous surfaces such as water (figure 26).

Figure 26: Bright line for detector 337 tracked across the ocean surface.

These features often seemed to increase in clarity close to coastlines or near clouds, but disappeared very rapidly as the radiance values increased.

Overall the validation confirmed that the persistent residuals align with observed striping in the imagery. We would recommend for a rapid assessment to use bands 20 and 21 over water, to confirm the validity of our findings.

As we gathered evidence we noted that the feature presence, although it could be found in many cases, did not behave in the way we expected based on the non-linearity plots we had generated. Even with similar radiance values, obvious stripes would simply disappear as the radiance increased slightly, such as that in figure 27. This was especially true of the longer wavelength bands (Band 10 onwards).

In figure 27, a dark stripe is seen next to the cross-hairs. This is over water with a radiance of around 3W. As the stripe moves towards the cloud, it fades and disappears rapidly, even though the area in which it fades is still around 4W. This sudden disappearance was not expected to happen so rapidly. Given that if it is a non-linear effect with radiance, we would expect it to only disappear as it entered the cloud based on the non-linearity plot we have for this detector.

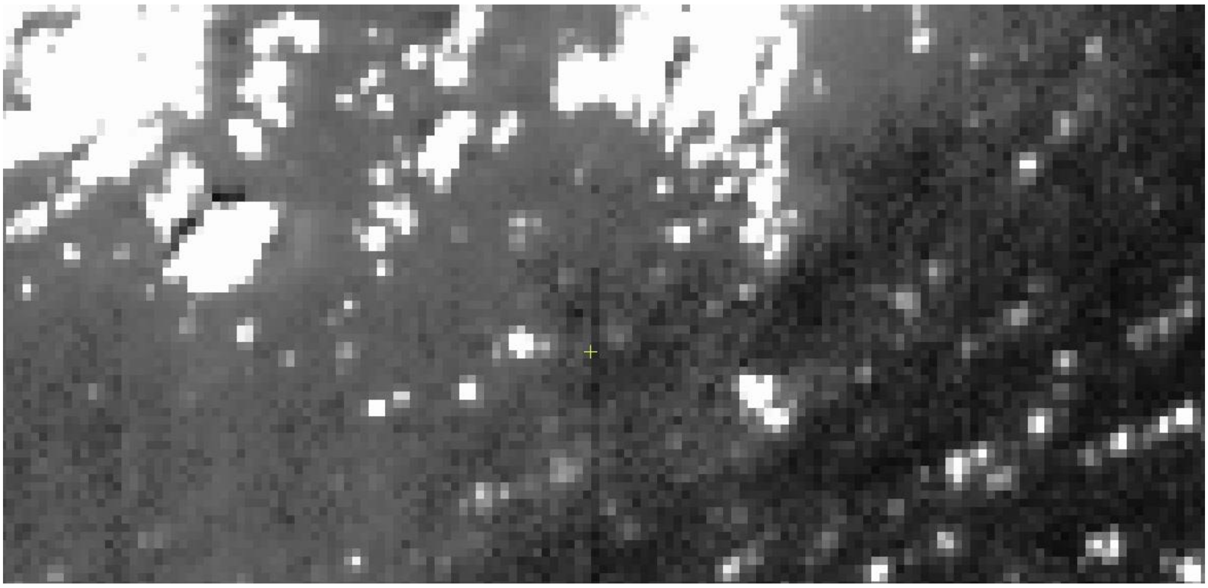

Figure 27: Band 21 feature, which just fades as we move towards clouds,

This rapid change made us consider that perhaps there are other effects we need to take into account or something we had overlooked in the non-linearity determination.

So the validation exercise was conclusive, in that where the relative gain algorithm finds residuals we find the presence of striping in the imagery.

Future work

It is necessary to understand the relative gain model applied in April 2019. It did have an impact on the residuals observed in bands 1 to 4 of the imagery (based on SNR analyses and post change relative gain analyses) but had no impact on band 13 and a potential negative impact on longer wavelength bands, such as Band 20 where residuals increased in magnitude after the model was applied.

5. Results - Level 1 OLCI Non-Linearity

The non-linearity is a simple extension of the relative gain analysis, but this time tying the observed persistent residual to the average radiance of the image being examined. So essentially using the same relative gain information but adding ancillary information on the scene brightness.

The origin of the approach was to try to explain why persistent residuals exist; that is relative gain variations of type 2 are present as discussed in Section 4 of this report. The only way to determine if there is non-linear behaviour is to examine each individual detector pair response to derive non-linearity relationships between the two.

So for every detector of every band for all the cameras, we have a corresponding non-linearity curve as shown in figure 28.

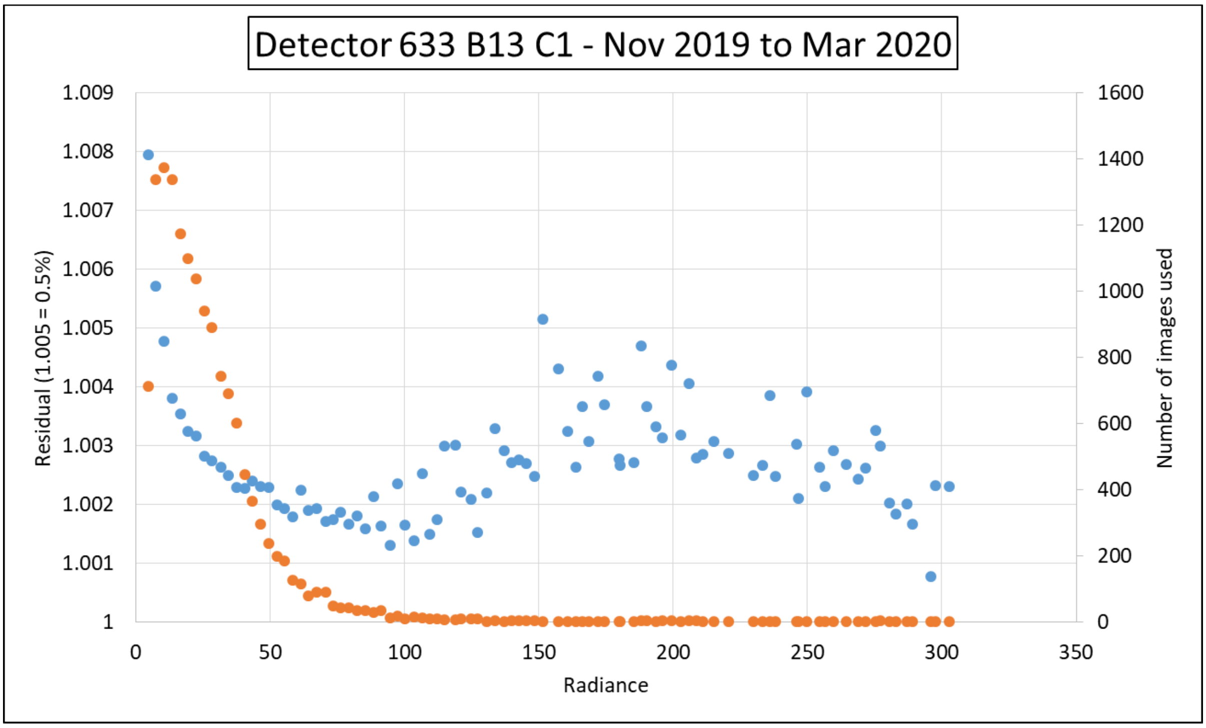

Figure 28: Residual variations due to non-linearity (blue) and number of images processed (orange) for detector 633 of Band 13, Camera 1.

The blue dots show the response. The x-axis is the radiance, so the scene brightness, the y-axis is the associated persistent residual where 1.005 = 0.5%. As we can see there is structure in the response for this example (Band 13). The orange dots show the number of images that have been averaged to derive that single point of the blue dots.

We can see two relationships for this band, a rather sharp variation at low radiance values and then a gentler meandering with much more scatter (given the lower number of images contributing). So a quite variable response.

For other spectral bands we see different responses (figures 29, 30 and 31), which seem to show a particular pattern of behaviour as we move from shorter wavelengths to longer wavelengths.

In figure 29 we see the band 1 response, the magnitudes of the residuals are very small and the behaviour shows an almost linear ramp up to the perfect calibration line at high radiances. So a simple relationship, but not additive in nature.

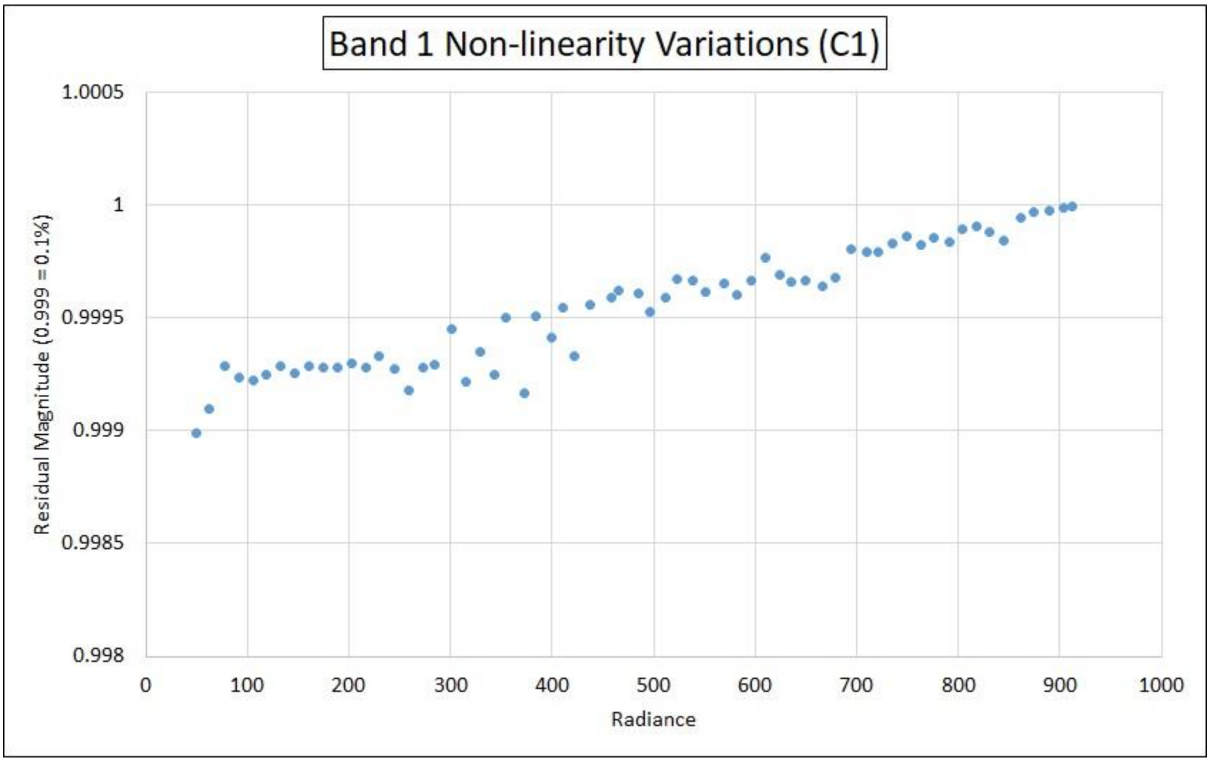

Figure 29: Band 1 non-linearity response.

Moving to longer wavelengths (Band 10, figure 30) we still have the simple response seen in figure 29, but also a very large change in the residual size at low radiances, which we believe could be an additive term (bias error). The values actually extend to lower residual magnitudes, but have been truncated so we can see more detail of the main plot.

The additive component tends to add an asymptotic effect, in which close to zero radiance we see large percentage deviations, as the signal level is small and the additive term is in relative terms much larger. As we move to higher radiances the additive term becomes correspondingly smaller and the line would, if it was purely additive, converge on the "1" line of no calibration difference.

However, in this case, in Band 10 there is an additional component which has induced a slope variation with a deviation at lower radiances (40W) of 0.08% of the signal level, converging to the no calibration difference line at a radiance of 500W.

In terms of image quality we would expect to see striping effects at very low radiance values < 10W in this example a bright stripe. As the radiance increases this would disappear and give a faint dark stripe that might be seen over homogeneous surfaces. As the radiance increases further this would disappear totally, so that over bright targets such as snow-fields or the on-board diffuser, no effects would be seen.

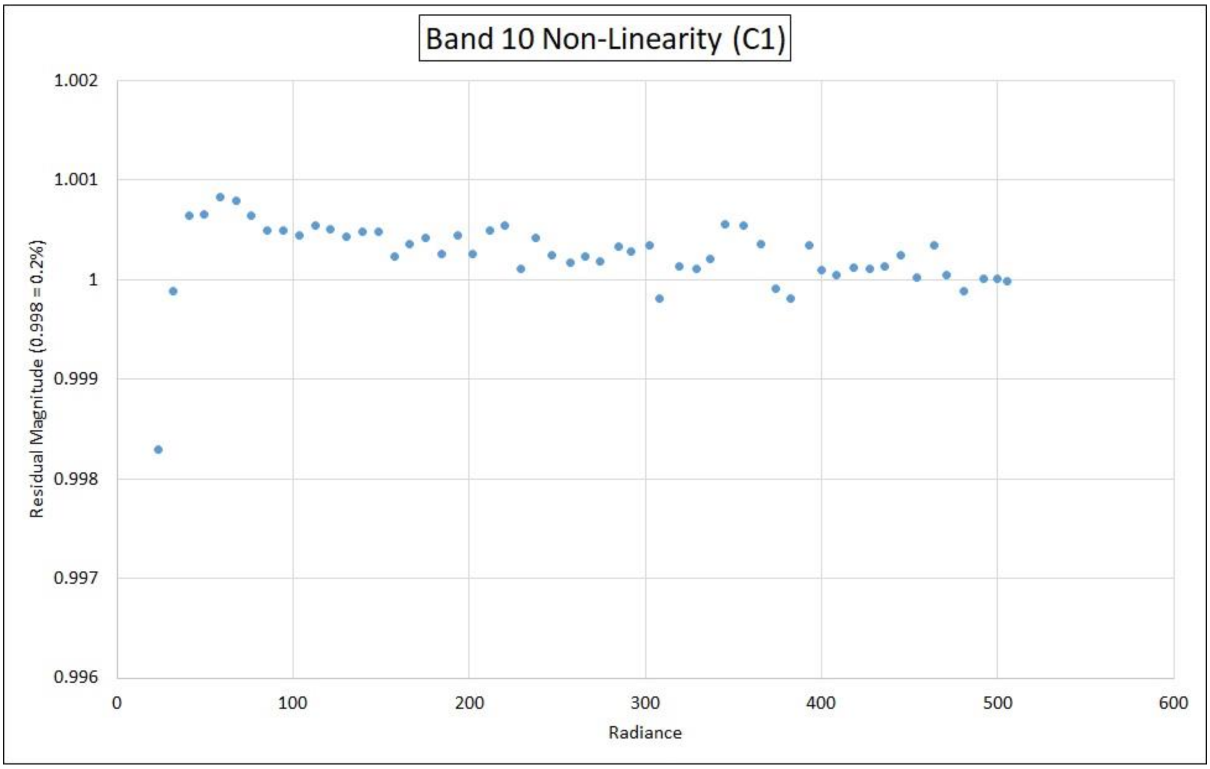

Figure 30: Band 10 non-linearity response.

If we go to even longer wavelengths, such as Band 20 in figure 31 we see a very common form for the data in which the non-linearity curve is dominated by what we believe are additive terms.

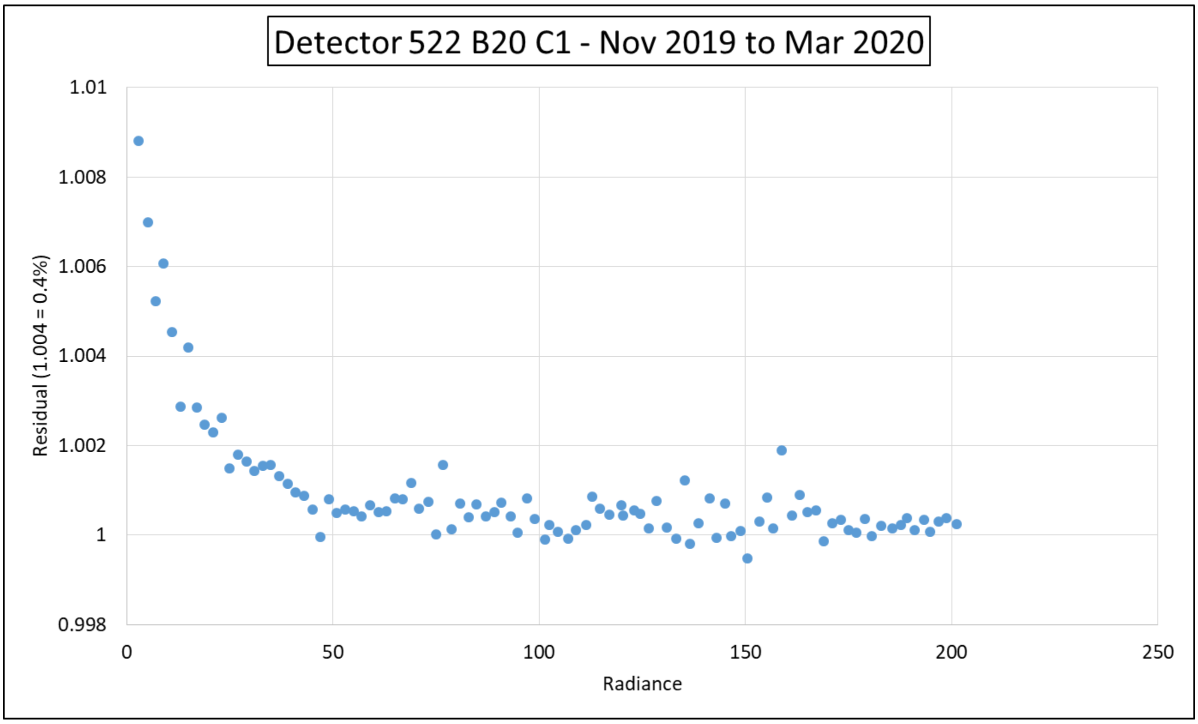

Figure 31: Additive bias errors cause a typical non-linearity curve, converging on the “1” line.

For Band 20 in figure 31, we see a sharp change at very low radiance that converges reasonably rapidly to the perfect calibration line. The behaviour looks almost asymptotic and can be modelled almost perfectly with an additive term error. Similar behaviour has been seen in Band 21.

The impression one gets from analysing the data is that the lower band numbers (short wavelengths) have multiplicative non-linear effects, as we move to longer wavelengths we see mixed behaviour (such as Band 13 in figure 28) and then at very long wavelengths, almost entirely additive behaviour.

It is not clear why this should be the case, as the dark shutter data is used every few weeks during the calibration cycle, so additive residuals should be minimal. However, there have been some discussions with ESA representatives of stray light issues in the longer wavelength bands.

If the stray light was a function of wavelength, then we would expect to see a more additive term dominated behaviour in other bands (band 19, band 18 etc.) and this seems to be the case (figure 32). In figure 32 we can see the moderate drop off, it was noted that there is more scatter in the low radiance region in Bands 18 and 19, but all detectors sampled showed more or less the same behaviour of higher residual values at low radiance increasing in a pattern which seems to reflect some additive component, although less sharp a change than that seen in bands 20 and 21.

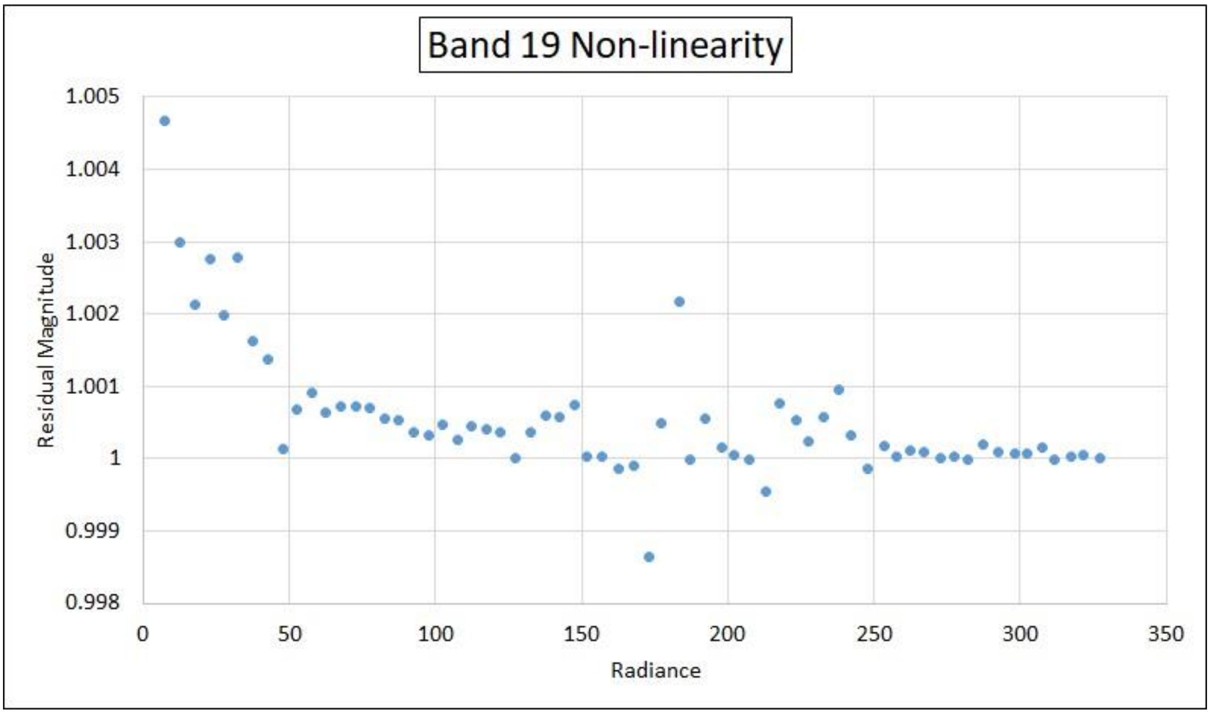

Figure 32: Band 19 non-linearity output

Given the results we could in theory formulate a correction between any detector pair. However, the findings in the relative gain analysis suggest that if stray light is an issue it may complicate some of the corrections. This whole area has emerged at the end of the project, so it is difficult to determine exactly the cause of these different non-linear effects we are seeing.

Given the variation in the non-linearity response as we move from the shorter wavelength bands where a more multiplicative regime seems to be present, towards band 13 where we see elements of additive and multiplicative and finally towards band 20 and 21 where we see an almost pure additive regime, it would suggest that if (and this is a big if) the stray light is the issue, then it affects data from at least Band 10 to Band 21.

Validation:

The only validation available is by analysis of imagery. These features cannot be seen using the on-board diffuser. However, as discussed in the relative gain validation section, we see some unusual behaviours which do not match our expectations. For example we see features in Band 21 in which a single dark stripe can fade rapidly and disappear as we approach cloud. This is in many ways what we expect to see, but in this case the fading is too rapid, with only a small change in radiance. We would expect to see a more gradual decline in values. So although we can see the fading, which proves that the result is dependent on the radiance of the scene, we still need to explain the rapidity of the change.

We considered possible sources of why it was different from expectations. We now believe it is due to the way the relative gain algorithm operates.

Imagine we have a scene in which the top 25% of the scene is covered with a bright layer, cloud for example, with an average brightness of 70W and with no relative gain residual and that the other 75% is covered by water with a radiance value of 5W and a 2% relative gain residual. The algorithm calculates the average residual, say for example we use the mean (there are options in the algorithm to choose different measures), which will be affected in part by the different contributions and also the algorithm calculates the average radiance in the column being compared.

We will end up with an average radiance of 21.25W and an average feature size of 1.5%. One entirely over water will give an average radiance of 4W and feature size of 2%. So depending on the mix of materials in each scene we will get lots of scatter, which we did observe in the original raw data with low/high radiance mixes (figure 33 from the January 2020 monthly report).

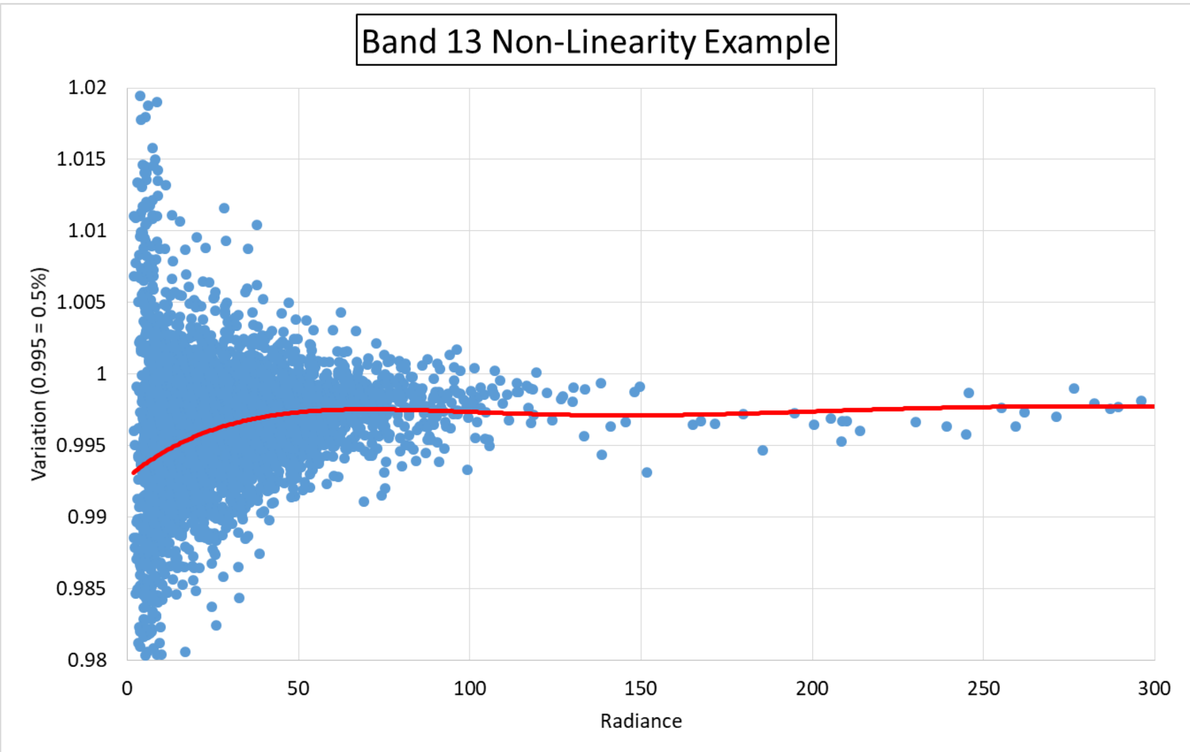

Figure 33: Data cloud showing scatter in residual values due to mixture of different surfaces

This mixing of surface components can adversely affect the shape of the non-linearity curve, producing a curve in which the convergence to the "1" line is slower. The impact this has on the validation, is that we expect a very gradual disappearance of a feature based on our non-linearity curve. In figure 34 we created a simple additive model with a 0.1W additive component for radiances up to 70W.

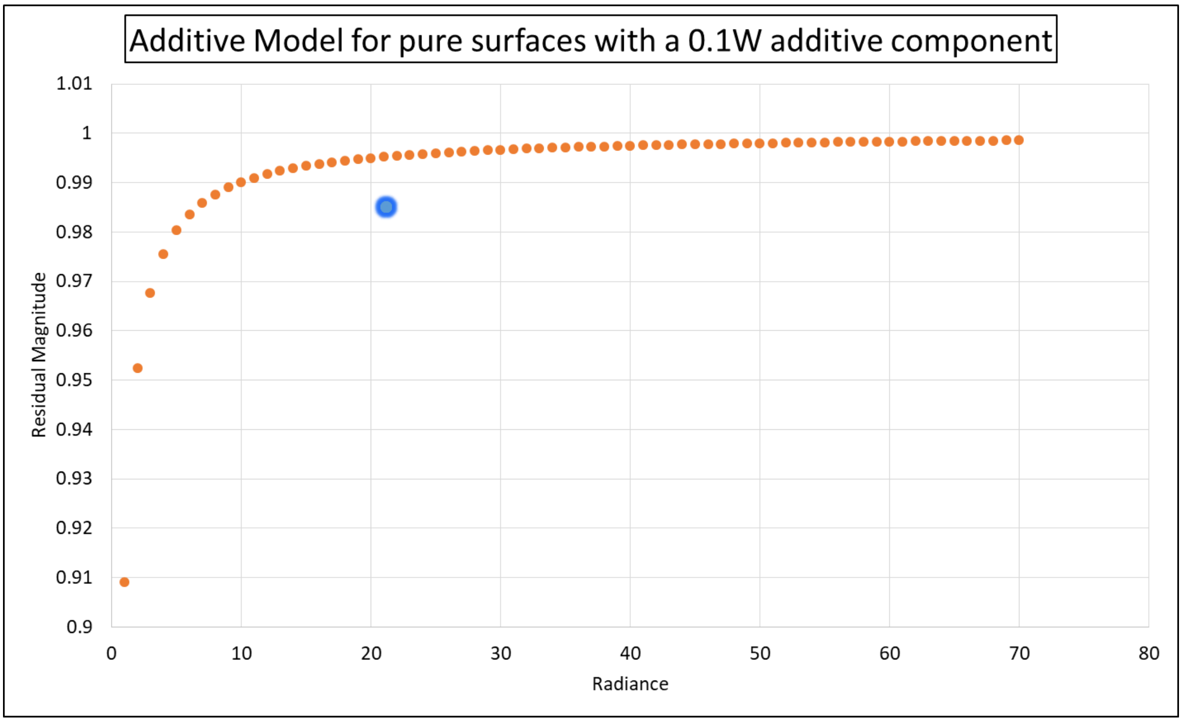

Figure 34: Additive model, the blue dot is the expected set of values for our mixed surface

As can be see it shows the asymptotic qualities we noted for Bands 20 and 21. However, we also calculated what would happen if we had a mixed surface for two components of 75% of material at 5W and 25% of material at 70W, the blue dot in the plot.

If we think about this, the conclusion would be that the non-linearity plot, when enough samples are included, will follow a line that is typical of the mixture of surface components normally captured and not that of specific materials of a particular radiance value. This will tend to produce a more gradual, less sharp curve that would deviate from one showing a pure additive term.

So it means we are overestimating the feature depth at higher radiances. This will apply to all spectral bands. However, in cases such as Band 1 where the additive term is less significant and the difference between different radiance values is small then it should be less of an issue.

The difference is small in most cases, especially for the shorter wavelength bands where the difference in residual magnitude between different radiance values is very small. However, as we move towards the longer wavelength bands, where the additive component increases, we will see larger deviations, especially in the very low radiance values. This effect would explain why we see features disappear more rapidly than the non-linearity plots suggest when validating the results.

The only way to deal with this is to modify the methodology to collect the non-linearity information, something we examined earlier in the project but abandoned due to the very large storage requirements to capture so much data.

Future Work

I would recommend some additional research to gain a better understanding of the stray light correction for OLCI. I believe that most of the larger residuals we are seeing in the data are related to the very low radiance regions and that the major component is an additive term. Unfortunately many studies over Oceans have very low radiance values, so the corrections could be significant in improving data quality.

Better mappings of the non-linearity could be obtained by using radiance values accumulated at the pixel level, rather than the scene or granule level. However, this would require re-writing the non-linearity code and based on work earlier in the project, there would be a big impact on storage requirements if we wish to keep flexibility in terms of selecting date ranges for analysis, as we would then need to store all the raw outputs.

If we are happy to aggregate all the results obtained, then the storage requirements are reduced considerably, but then the results will always be aggregated and sub-sets could only be extracted by re-running all the images a second time. This is why the simpler scene / granule based approach was used, as the storage and retrieval options were better and the results assuming a multiplicative model (the original model envisaged) would have produced reasonably accurate results.

However, it’s now clear that there is an additive component a bias error of some kind which affects the longer wavelength bands considerably more than the shorter wavelengths. It would be wise to consider carrying out such an experiment, using a pixel based estimate for the longer wavelength bands at some point in the future.

Note that if ESA successfully removes the additive component then this problem will largely disappear and the much easier correction for the more linear behaviours can be applied.

6. Results - SLSTR Level 1 Processing and Analysis

There is only a single algorithm for the Level 1 processing of the SLSTR data, which is the Signal to Noise Ratio (SNR) estimation. The algorithm used is exactly the same as used for the OLCI analysis, using the standard moving window algorithm with a 7x7 window size.

Initially it proved quite difficult to carry out the SLSTR analysis due to the limited number of images producing a very sparse data cloud especially at high radiance values. The early data strongly suggested that the SLSTR data was not shot noise limited, with a linear ramp at low radiances. However after accumulating enough data, we were able to define the more typical shape we expected. There is still a strong linear component, but the rest of the curve is fairly typical (figure 35).

In figure 35 we can see that the SNR data clouds have a clean upper boundary and that in this particular case for Band 1, that the Oblique view data of the A stripe has much higher SNR than the Nadir view.

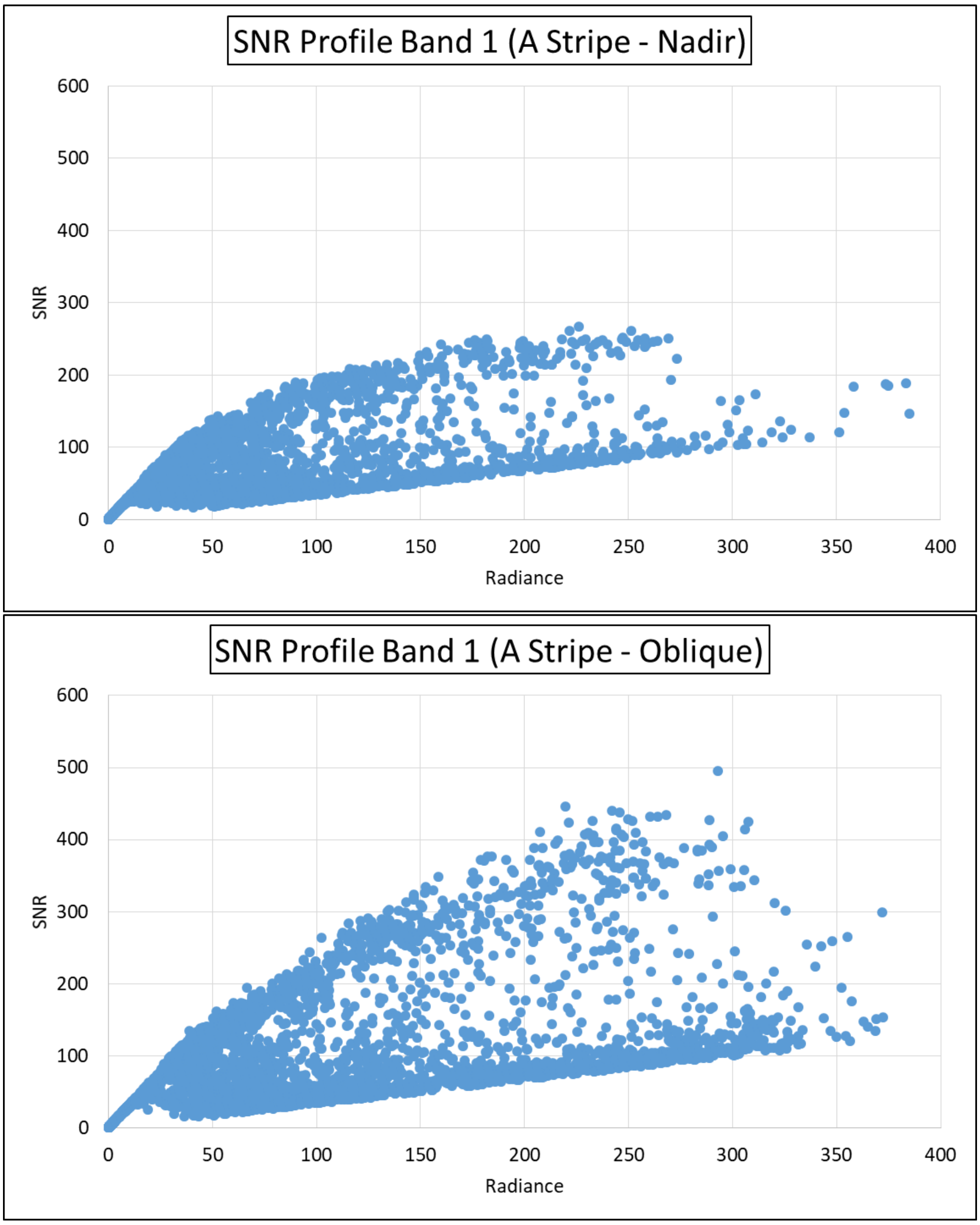

Figure 35: SNR for Band 1 (A Stripe) for nadir (top) and oblique (bottom)

If we plot the data slightly differently, this time signal against noise, you can see the effects of this linear behaviour at lower radiance values (figure 36). Linear behaviour in SNR plots show the same noise for increasing signal. So a signal of 100 units with a noise of 1 has an SNR of 100-1, double the signal and keep the noise as 1 we go to 200-1, so this produces a nice line in our SNR plot.

The curving behaviour shows that as the signal increases, we are also seeing an increase in the noise. Looking at a plot of signal against noise we can see that there is a plateau of almost unchanging noise between 10 and 30W radiance on the x-axis (figure 36). So the noise increases, then plateaus for this 20W range and then continues to increase.

This type of behaviour has never been observed in any other instrument. The cause of this behaviour is unknown at this time.

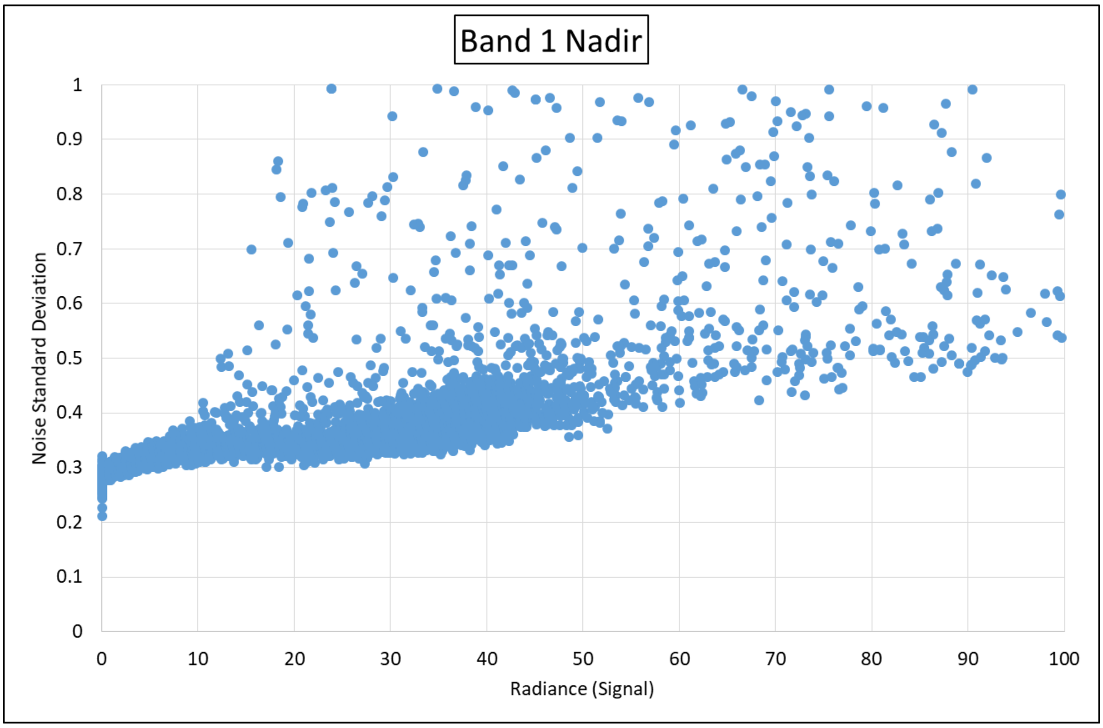

Figure 36: Signal against noise plot for Band 1 Nadir showing plateau of noise levels

EOSense has examined many different instruments and the normal signal to noise plots show a rapid increase in noise initially which then turns into a more linear form, such as the example for OLCI in figure 37.

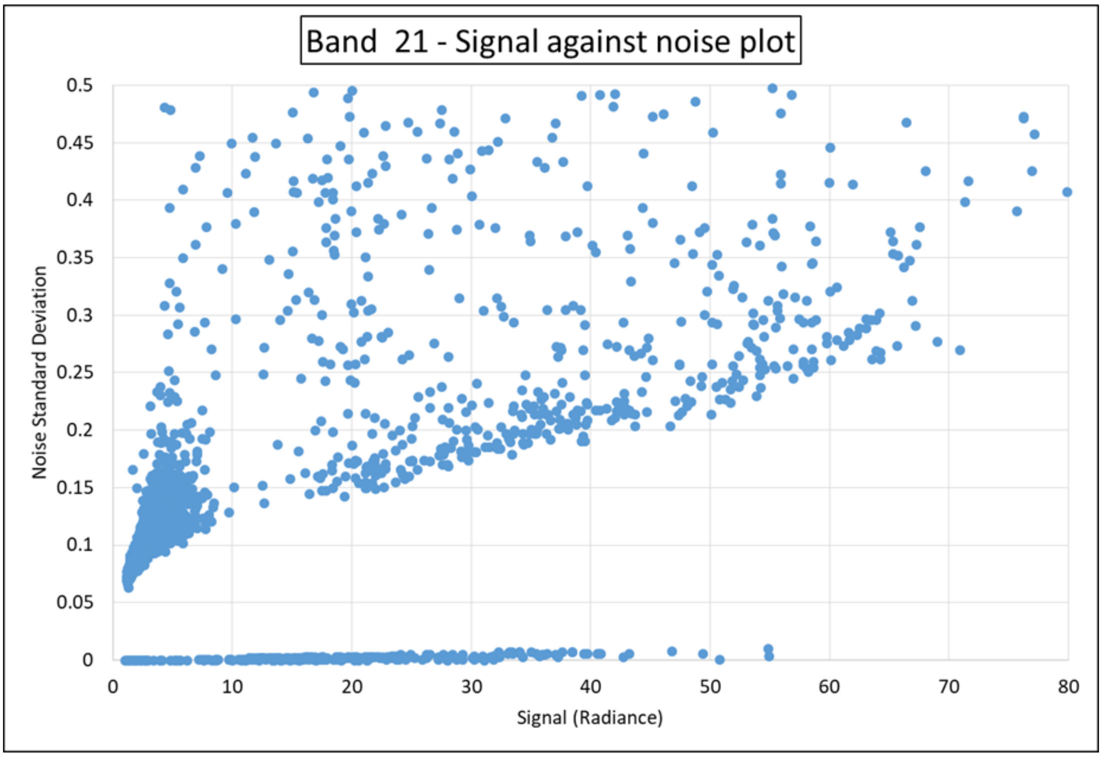

Figure 37: Signal to Noise plot for OLCI band 21.

The SNR was determined for all the reflectance channels for oblique and nadir and for A and B stripes in cases where both stripes were present (bands 4, 5 and 6). The results are summarised in table 4.

Band	View	Stripe	Target Radiance	SNR
1	Nadir	A	100W	200
1	Oblique	A	100W	250
2	Nadir	A	100W	285
2	Oblique	A	100W	325
3	Nadir	A	100W	400
3	Oblique	A	100W	395
4	Nadir	A	4W	40
4	Oblique	A	4W	35
4	Nadir	B	4W	35
4	Oblique	B	4W	30
5	Nadir	A	2.5W	90
5	Oblique	A	2.5W	50
5	Nadir	B	2.5W	100
5	Oblique	B	2.5W	65
6	Nadir	A	2.5W	65
6	Oblique	A	2.5W	85
6	Nadir	B	2.5W	100
6	Oblique	B	2.5W	75

Table 4: SNR values for SLSTR Level 1 data for all bands, views and stripes.

We had expected to see some pattern in the results. The result is based on using data with excluded pixels and cosmetic infilling, so we expected generally better values in the oblique measurements given the pixel replication for the cosmetic infilling. However, analysing all the results there were roughly twice as many results with the nadir showing higher SNR than the oblique. Of the first three bands, Band 3 had the highest SNR and band 1 the lowest and on average B stripe data had better values for bands 5 and 6, while Band 4 has better A stripe results but only marginally.

So no pattern emerged regarding oblique and nadir data or A and B stripes.

Validation

No validation has been performed at this time. Although the preferred approach is documented in the Product Validation Plan, again it can be compared against literature. However we need to make sure that any estimates of SNR in the literature are directly comparable to those derived from the heterogeneous images, so that a direct comparison can be made.

If differences are found between the SNR in the literature at specific target radiances and those derived from heterogeneous images, we may need to consider using other approaches, such as homogeneous surfaces (such as snow scenes) and a careful evaluation of the pre-launch estimations.

Future Work

It would be an interesting study to compare the data clouds produced by using the cosmetically filled data product and that with the reconstituted original imagery to determine the impact on the SNR estimation. We expect the difference to be small as the pixels are adjusted using a nearest neighbour approach rather than any form of more complex convolution.

7. Summary Discussion

In this section we will summarise the discussion of the results and frame a series of questions. At the beginning of the study one of the aims was to evaluate if the use of the heterogeneous images could **help validate the results derived from on-board systems**.

The results if anything, have pointed to differences rather than fully reinforcing the results obtained from on-board systems. Some items are yet to be fully evaluated, such as the Level 0 results related to calibrator drift and absolute calibration drift. Once sufficient data is collected, the comparison can be made and the full effectiveness determined.

For the other items such as Level 1 SNR for OLCI, the surprising conclusion is that the results from the heterogeneous images show all the same relative information, such as which Camera system is noisier and which band is noisier, fully recreating the SNR profile across the 21 bands. However, there is a disparity in the absolute SNR obtained. This in part may be due to the way SNR is determined on-board and differences from the methodology used for the heterogeneous images. However, even taking possible differences such as column based estimates against area based estimates into account, we cannot close the gap between the two sets of measurements. The interesting finding from our point of view is that both the limited pre-launch data we have and the homogeneous scene data from snow scenes gives similar estimates to the heterogeneous images, suggesting the SNR estimates on-board are overestimating the SNR.

For relative gain, our methods highlight differences. The relative gain values in the data according to results based on the on-board diffuser are so small they should not be visible in the imagery. Just a cursory glance at a band 20 or band 21 images over the ocean and it will become clear that there are strongly visible features related to relative gain.

The issue in this case seems to revolve around the lack of information collected about low radiance targets or even earth background radiance targets. Features are present, usually quite small, but large enough to leave visible effects. Evidence from the non-linearity plots suggest that these features are very prominent in the low radiance regions and tend to disappear as we move towards bright targets.

This in one way validates that the methodology based on the on-board diffuser is working correctly, in that it sees no features. However, there is no available methodology to look outside the radiance range of the on-board diffuser or use of the shutter for dark images, hence the features we see are not captured. There is a strong suggestion in the results that an additive term error is responsible for most of the larger deviations, mainly affecting very low radiance values and longer wavelength bands. However, there are some significant, but small, features in all bands that suggest other calibration errors are present that affect low to mid-range radiance values, which could be corrected.

In summary, the results have confirmed some of the results obtained from the on-board systems, such as relative SNR differences, and that at very bright radiances the on-board diffuser and our results would coincide in confirming very low relative gain residuals. However, the big advantage seems to be more in **highlighting differences and providing insight into possible causes of the observed differences**.

Being able to highlight these differences, helps point to what is causing these differences and potentially how to account for them. For example the relative gain estimation was originally envisaged for finding single detector events (our type 1 relative gain residuals), to highlight them rapidly rather than wait for the fortnightly calibration updates.

However, it soon became apparent that persistent residuals existed, that continued to be present through multiple calibration cycles and could not be seen by the on-board diffuser or identified using the dark shutter data. This required an explanation, which pointed to some non-linear behaviour, which generated a non-linearity algorithm as an offshoot from the relative gain algorithm, which by the nature of the results obtained pointed to an additive term error affecting the longer wavelength spectral bands of OLCI.

So by simply finding the features we are starting to uncover behaviour in the instrument that was not previously recorded or fully understood. So **uncovering new information** has been a major benefit of the study. Some items need further study, certainly understanding how the SNR is calculated using the diffuser measurements might help in determining why we see a disparity. For SLSTR the unusual signal to noise profile needs some explanation as it affects SNR estimates at lower radiance values over the ocean.

7.1. Future Activities

In this section we outline the sort of continuation activities required to maximise the amount of information that can be extracted using the methodologies.

7.1.1. Level 0

Ideally for the Level 0 studies, it would be best to continue running the algorithms and produce weekly summary values for an entire year. In the case of OLCI this profile could then be adjusted using both diffusers to get what would be more or less a robust annual curve.

There will be deviations in this curve which should be visible in multiple sensors, but it would require further work to determine exactly how the deviations could be removed. These deviations would be due to anomalies produced by clouds for example that affect the weekly average counts.

Once this curve is generated it can be the reference for further years. SLSTR would require an alternative approach, which could be based on using the OLCI data for bands which overlap.

7.1.2. Level 1

For Level 1 OLCI SNR, apart from trying to understand the causes of the disparity between the on-board measurements and the heterogeneous images, one option that has been discussed but not implemented was the further development of the column based SNR estimates. There are at least two or three different approaches that can be considered. These would help validate the window based solutions as the principle of operation will be different, by removing the spatial averaging element so should give marginally higher SNR values. Some are in early stages of development, but none are implemented at this time.

For Level 1 SLSTR SNR, the validation still needs to be performed. The only extra item of interest is to determine the effects of removing cosmetic pixels and replacing orphaned pixels and re-running the SNR estimations. We expect this will give marginal changes to the final result.

For Level 1 Relative Gain, the algorithm works very well in identifying the presence of relative gain differences not seen by the diffuser. The causes now seem to be clear, that non-linear behaviour is present due to a mixture of additive errors and multiplicative errors. Hence the focus should be on the Level 1 non-linearity algorithm. We discussed that the algorithm at present is using a rather coarse approach, but shows the major causes of the relative gain variations, primarily additive errors for longer wavelength bands, different from the shorter wavelength features observed.

It was also noted that the non-linearity methodology could be improved, requiring a larger output database, but sacrificing our ability to select date ranges of interest for analysis. The methodology will also be slower in operation as it is far more complex to implement and run. Some studies carried out in late 2019 and early 2020 explored the idea, but it soon became apparent that the database sizes if we wished to have date selection would become unmanageable very rapidly. However, as long as we just took the whole non-linearity database (which could be split manually for processing) we could keep the database size down to a manageable level.

Until we get an accurate relative gain model for each detector it would be difficult to move towards a set of correction factors to remove the observed effects. Any correction would require choosing a reference or set of reference detectors or choosing some optimisation strategy that minimised the amount of change we were making to the detectors.

For example, we could choose any detector as our reference and move outwards and correct from that detector. Some detectors would be further from ideal and would cause a larger ripple of changes to correct all other detectors. Others would cause a minimum of changes to surrounding detectors. The algorithm would have to be complex as it would use non-linearity inputs so it would also need to take the radiance into account for the correction. No attempt has been made at this time to scope out the work required to make this type of iterative correction.

7.1.3. Interface Elements.

The interface as it currently stands seems to be effective enough. It is semi-automated. The user starts the process and all files in the input folder are processed and added to the databases. Once complete the user places a new set of files to be processed and runs them through. So it requires the user to move files or create symbolic links to files and run the process.

It gives some flexibility on which files to use and to re-process if necessary, whereas a fully automated process would be taking the data from an external source as data arrived and the automated process would have to be stopped if re-processing were to take place.

The new analysis algorithms currently provide summary data for:

- OLCI SNR
- OLCI Relative Gain (by date selection)
- OLCI Non-linearity (by data selection)

The OLCI SNR could also be improved by using a data range. Also the initial relative gain and non-linearity analysis algorithms are not as efficient as they could be for the extraction and could be improved. It would be advisable to add additional algorithms to extract Level 0 summaries over specific dates and to be able to quickly select and plot non-linearity curves for any specified detector. However, none of these changes are essential and the core algorithms included are the main ones to speed up the key data extraction.

7.1.4. Additional Validation

For some algorithms it would be wise to use multiple sensors in either the processing and/or the validation. For example the level 0 counts data can have anomalies due to cloud cover for example. Using two sensors it would be possible to isolate these anomalies and remove them from the trends.

For much of this study we have had to resort to alternative methods of validation, due to the differences seen between our estimates of SNR and relative gain and those derived from the on-board diffuser. The process is very manual and could be improved by developing simple profiling tools, that took average deviations across homogeneous sub-samples to compare to the outputs from the relative gain algorithm.

7.1.5. Application to Other sensors

The methods demonstrated are applicable to a wide range of different sensors, both low and high spatial resolution. For relative gain and non-linearity, pushbroom sensors are the ideal choice, while for SNR, any data set can be used to get an assessment of whole image SNR.

8. Conclusions

Finally we can conclude the following from this study at this time:

Level 0 – OLCI

- The Level 0 data aggregated at the weekly level does show coherent patterns of behaviour over a five month period, which suggests we may be able to define a reference curve for both the calibrator drift and the absolute calibration drift. The data collection needs to continue to a minimum of the end of November 2020 to at least assess the methodology.
- Multiple sensors may need to be used to remove anomalous values in the weekly aggregates related to periods of unusual cloud cover for example.
- The approach looks very promising as an alternative/compliment to traditional calibration.

Level 0 – SLSTR

- Due to anomalies in the code provided, no SLSTR database data is available. These anomalies were corrected in February 2020 and we expect that data is or soon will be collected for this sensor. A full year of data is required for this sensor and this should overlap with the OLCI data collection of Level 0 data, so that the OLCI data can be a reference for the SLSTR data analysis.

Level 1 – OLCI

- Three algorithms were implemented. SNR, relative gain and non-linearity. All three gave very useful results which was part validated by the on-board devices, but also raised some important questions on the results currently obtained by the on-board devices and potential issues with either the methodologies being used or the sensor operation.
- The SNR values mirrored the on-board diffuser measurements, showing similar relative behaviour between bands and between cameras. However, there was a large disparity in the magnitude of the results obtained, with the heterogeneous image analysis giving SNR values that were tens of percent lower in most cases.
- The SNR values obtained were cross-checked against a single pre-launch source and snow scenes of homogeneous nature. These results were consistent and all were lower than the diffuser estimate. Details of the diffuser calculation of the SNR are required to try and identify why there is such a discrepancy.
- Part of the SNR discrepancy can be accounted for by the use of spatial statistics for the heterogeneous image analysis and the snow scene analysis, as if there are relative gain differences, these will tend to give lower SNR values. However, modelling suggests that this will at best only account for half of the difference observed and only for the first four spectral bands.
- The relative gain differences observed consist of two types, simple deviations from normal for single detectors that can be seen by both the heterogeneous imagery and the on-board diffuser. These are the type 1 residuals discussed and can be easily removed. However, we observed persistent residuals which have been present for at least two years which are not detected by the on-board diffuser and cannot be corrected. Validation using simple image analysis shows the observed features for Bands 20 and 21 are clearly visible over homogeneous water bodies.

- The relative gain effects are greatest in very low radiance regions especially in the longer wavelength bands and have the aspect of an additive (bias term error). For shorter wavelength bands there are residuals of lower magnitude which are not additive.
- Corrections of the relative gain model (such as that in April 2019) tend to reduce the relative gain effects at shorter wavelength bands, have no effect on middle wavelength bands of OLCI and can have a negative effect on the longer wavelength bands.
- There is a pattern in behaviour of the relative gain residuals which suggest small calibration errors at shorter wavelength bands, a mix of additive term and multiplicative terms in the mid wavelength range and almost pure additive in the longer wavelength range. The additive term effects have the biggest effect in defining the presence of large relative gain residuals.
- The non-linearity algorithm can be used to deduce the type of effect seen (multiplicative, mixed or additive) and provide some indication of the magnitude of the effect.
- The non-linearity algorithm in its current form seems to be over estimating the relative gain residual due to the way the algorithm operates. Using average values from a scene, instead of gathering statistics at the pixel level. Improvements to the algorithm would produce a more accurate non-linearity mapping that could be used to correct for these effects.
- These novel methods have given a great deal of insight into the limitations of current methods and some of the potential instrument issues generating these features that we can detect.

Level 1 – SLSTR

- The SNR algorithm produced very good results. There seems no pattern in which view gives the higher SNR. Over all the bands, two thirds are nadir view and one third oblique, which suggests that the cosmetic infill has little impact on the SNR estimation.
- The data without cosmetic infill should be also examined with the SNR algorithm to determine exactly the difference it makes in the SNR estimation.
- There is no specific difference between the A stripe SNR estimates and B stripe estimates for bands 4 to 6. In some cases the A stripe is better (band 4) and others the B stripe (bands 5 and 6). In terms of the first three bands, band 1 has the lowest SNR and band 3 the highest.
- There is some unusual behaviour observed in the signal against noise plot which at the moment cannot be explained. A plateau of noise is reached between 10W and 30W radiance range in which the noise is essentially constant. This type of effect should not be present in a shot noise limited system and this behaviour has not been observed before with other sensors.
- The generated data clouds provide a lot of detail which suggests that this sensor does not have a simple response and that its noise profile is by no means a standard shot noise limited curve. No other methodology can derive this information with this level of detail.