EOSense

STATISTICAL DEGRADATION MODEL FOR OPTICAL SENSORS: Algorithm Theoretical Basis Document (Draft)

Project No. 18

EOSense-018-011

V2.0, 31/05/2020

Authors: S. Mackin and J.Settle

Contents

1. Introduction

After this introductory section, this document is divided into three parts. The first part covers algorithms that are using Level 1B data products as input, while the second part covers algorithms that use Level 0 data. The final part covers topics related to the software implementation and evolution of this document.

Given the research driven nature of the project we expect this document to be a "living" document, that is, we expect it to change as the project progresses. As new avenues are explored we will add additional algorithms and as some areas prove fruitless we will remove them from the document. The aim will be to produce a full algorithm theoretical basis document that covers all aspects of the final set of algorithms implemented at Eumetsat.

1.1 Fundamental Approach and Overview of the Scientific/Engineering Outcome

In this document we describe the algorithms which together meet the set of requirements defined in the Requirements Baseline (EOSense-018-006, latest issue) that specifies additional calibration and data quality measures to assess the calibration of the Sentinel-3 SLSTR and OLCI imagers.

The underpinning idea for all the algorithms discussed, is that we can derive information on the sensor calibration and data quality from the imagery collected during normal operations. This information can be useful as a supplement to on-board systems, in other cases an alternative form of vicarious calibration and in some cases providing a unique source of information on issues that cannot be determined easily using on-board systems or standard methods of vicarious calibration.

Essentially we are using statistical measures of various forms to derive the information we require from the image data. In the case of SNR measurements it can be a histogram of the noise estimates from a moving window approach across a heterogeneous image. For relative gain it is the estimate in a calibration shift based on a distribution of ratio values between highly correlated neighbouring pixels. These two examples show single image estimates. For other operations we need a lot of data to derive a stable measure of the parameter of interest, such as calibration drift, in this case billions of data points provide a stable assessment of change from one week or one month to the next, with some key underlying assumptions about the stability of the earth's albedo.

1.2 Satellite Instrument Description

The OLCI instrument is the successor to ENVISAT MERIS with additional spectral channels, different camera arrangements and simplified on-board processing. It is a push-broom instrument with five camera modules sharing the field of view. The field of view of the five cameras is arranged in a fan-shaped configuration and each camera has an individual field of view of 14.2° and a 0.6° overlap with its neighbours. The whole field of view is shifted across track by 12.6° away from the sun to minimise the impact of sun glint. The spectral range is 400 nm to 1020 nm.

Calibration of all OLCI measurements are made via a calibration assembly of a similar design to MERIS that includes a mechanical rotating table. Either a direct view of the Earth (for imaging mode) or one of several calibration targets may be selected by rotating the table: a dark shutter plate (for dark current calibration), a primary Polytetrafluoroethylene (PTFE) calibration diffuser (viewed every 2 weeks for radiometric calibration), a redundant PTFE calibration diffuser (viewed every 3 months to determine degradation of the primary diffuser due to solar exposure) or an erbium doped 'pink' diffuser plate for spectral calibration. During the calibration sequence, a selected diffuser plate is moved into the instrument Field of View (FOV) and illuminated by the sun so that all five cameras can be calibrated at the same time. Characterisation of diffuser ageing is determined through on-ground processing using the two OLCI diffusers in synergy.

For more information, refer to the OLCI technical guide at:

https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-olci

The SLSTR instrument mission maintains continuity with the (A)ATSR series of instruments. The design incorporated the basic functionality of AATSR, with the addition of some new, more advanced, features. These include a wider swath, new channels (including two channels dedicated to fire detection), and higher resolution in some channels. The spectral range is 0.55 micrometres to 12 micrometres, though the spectral region of interest in this study is the VIS/NIR, with possible extension to the SWIR. The calibration scheme for the short-wave, near infra-red, and visible channels is based on a diffuse calibration (VISCAL) target of accurately known reflectance which is illuminated by the sun over a short segment of the orbit, and which is intersected by the instrument scans. The black bodies (for TIR calibration) provide a dark reference.

For more information, refer to the SLSTR technical guide at:

https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-slstr/instrument

Calibration and data quality activities are performed by the Sentinel-3 MPC, and the results documented in the cyclic reports at:

https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-olci/data-quality-reports

And:

https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-slstr/data-quality-reports

2. Part 1 - Level 1 Processing

Three specific development areas have been identified that use Level 1 data in the analysis, these are,

- Relative Gain
- Non-linearity
- SNR

The first two are intimately related, persistent relative gain residuals seem strongly associated with non-linearity effects between neighbouring detectors, these first two are only relevant to pushbroom type instruments therefore will only be applied to OLCI. The final area, the SNR evaluation is a separate algorithm based on a different approach and is equally applicable to both SLSTR (reflectance channels) and OLCI sensors.

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

2.1 Relative Gain

When a detector array is manufactured each detector in the substrate has slightly different behaviour, including

- Different bias values when there is no signal
- Some non-linearity in response
- Different overall response to the same signal level (gain values)

So to get a stripe free image from a group of detectors in a linear array we need to equalise all the detectors, so we get the same response to the same radiant energy on the detector surface. One of the steps of the calibration process is to determine the bias and gain values and correct for any non-linearity in response to avoid striping.

A raw image from Landsat 8 is shown as an example in figure 1, showing uncorrected data.

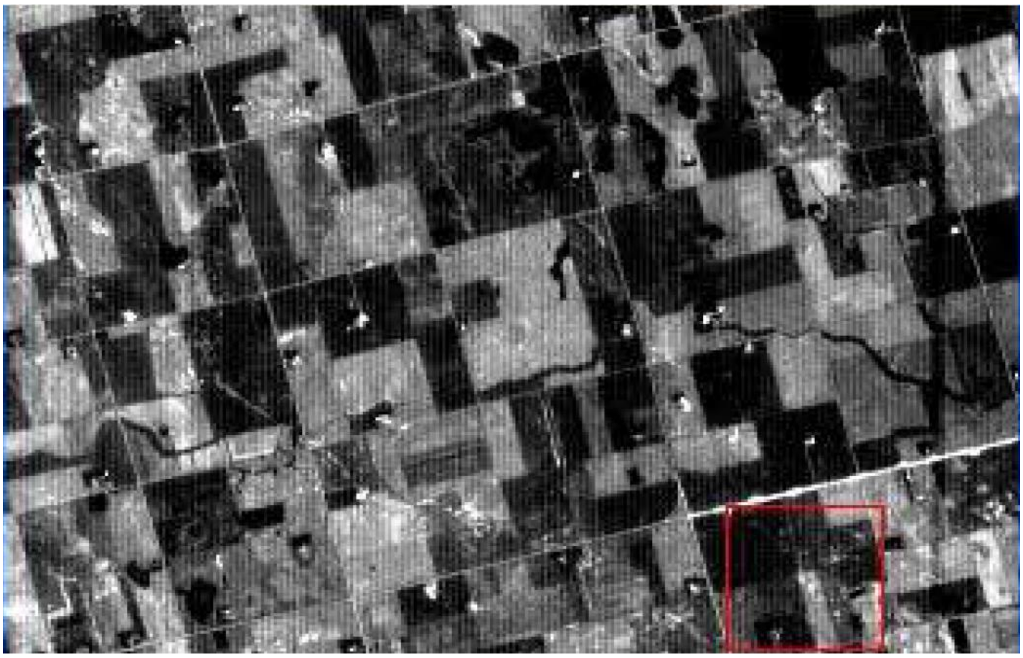


Figure 1: Landsat 8 image showing striping present due to detector to detector difference before calibration has taken place.

Once calibrated these detector to detector differences in figure 1 and hence the striping in pushbroom instruments disappear. Imagine if the image in figure 1 was a snow-field, with the same amount of radiative energy from each pixel of the scene. If we have performed the relative gain calibration correctly, no matter where we look in the image, all the values will be the same and no striping will be present.

The relative gain equalisation can be performed in many different ways, it can be performed on the ground using an integrating sphere or other light source that illuminates the whole of the detector array evenly and it can be performed in space using either an on-board diffuser illuminated by the sun and regularly viewed (every few weeks for Sentinel-3) or by using vicarious ground targets such as snow-fields. In all cases we use a homogeneous target to derive our measure of a uniform illumination for the correction.

2.1.1 Our Approach

The approach developed by EOSense is somewhat different, in that we can determine changes with a much higher temporal resolution than a daily visit to a snow-field (in the right season) or a monthly determination from an on-board calibrator. This is achieved by using all data as the source of our information, so single **heterogeneous** images, as soon as they are downloaded can be used to derive information on the relative gain of the sensor.

What do we mean by a heterogeneous image? We mean any image, not just specific calibration images (figure 2).

Figure 2: All images can be used in the analysis including very heterogeneous fields and urban areas

The relative gain algorithm can be applied in different ways to different data products. It could in theory be applied to Level 0 data to derive the relative gain curve in the same manner as we use flat-field targets on the earth or a diffuser on-board.

The problem with Level 0 analyses, is that we (EOSense) use heterogeneous images, we find gradients in brightness across our target area that introduce variability in our determination of the relative gain curve, which is not found with homogeneous surfaces or on-board diffusers. We can recover the higher frequency part of the Level 0 curve, which is not so closely related to surface variations, but it would require multiple images to average away the lower frequency variations induced by gradients across a heterogeneous surface to get a good example relative gain curve from Level 0 data.

There is an alternative to using Level 0 data, assuming that the initial calibration has been performed using either an on-board system diffuser or flat-field on the earth, we can use the Level 1 data produced and look for deviations away from a perfect calibration. If the calibration is perfect we should see no variation from one detector to the next as we go across the image, so in the case of pushbroom systems no residual striping effects, only very low level random variations that are not consistent from one image to the next should be present. This second approach using Sentinel-3 Level 1 data is the approach chosen for the analysis of the relative gain.

2.1.2 Principles

The basic idea we are exploring is that neighbouring detectors may not have the same response to same incoming radiant energy. Imagine we have a perfectly calibrated satellite and we are looking at a perfectly homogeneous scene where every pixel reflects exactly the same amount of energy. If we ratioed each pixel in column one with the neighbouring pixel in column two, we would get an exact ratio value of one for every pixel pair down the column.

Now let us add some reality, we add Gaussian distributed noise. Now we find the ratios vary a little around one, with the mean value of the noise histogram exactly on one and a standard deviation which is essentially the noise standard deviation of the imager for that brightness target.

Now let us add a calibration shift of the second column detector, making it 0.1% lower in signal level and then do our ratio map again. We have the same noise levels and we get the same histogram, but we see that the mean of our histogram has shifted to higher values to 1.001.

So essentially we are mapping the shift of this peak value. So let's now add more complexity, let's use a heterogeneous image instead of a homogeneous image. Now although our adjoining columns are very highly correlated, the correlation will tend to suffer, especially over very variable scenes, such as urban areas. So what happens to our histogram?

In this case, we have large ratio values pushed to the wings of our histogram, where we have sudden brightness changes between columns, from dark shadow to bright building or vice versa, we will also see small variations related to surface differences producing intermediate values, stretching out our histogram, making it broader and less well-defined. However, even with that, we have a lot of pixels where the material is similar enough that we are seeing the effect of the small calibration change.

The result is normally a more ragged distribution, in which we can exclude all the larger values as we know they are not due to calibration changes. In fact we could just try and focus on the peak position and using the peak shift we can estimate the relative gain change that has taken place.

However, we did mention reality, the problems we can find are,

- There are too few data points to define a really good histogram (especially over totally urban areas), so no easy way to determine peak value.
- The data is quantised with a large central peak that does not move (seen in Sentinel-2). However the secondary peaks at the next quantisation interval often show an asymmetry which is related to the calibration change (figure 3).
- In reality surfaces have cross-track gradients which affect (for a single image) the derivation of the relative gain terms.
- Although brightness effects are reduced we can see some structure in the initially extracted residuals that mask our attempts at deriving the correction factors we require to determine the true relative gain variations.

The basic principles of deriving the relationship are in themselves simple, the more difficult part is the interpretation of the corresponding histogram and the extraction of the correction factors over different surfaces which are at times very heterogeneous.

2.1.3 Step by Step

Step 1 – Creating the histograms

The initial step is to create a histogram of the ratio relationships between each pixel pair for each column pair. We work in log values either using all the data, or just data that has ratio values between 0.95 and 1.05. This second form eliminates the extreme values we discussed in the previous section, which we find in very heterogeneous areas such as urban areas. The two histograms produced using either all or part of the data are generally quite similar, which supports the general principle that the peak of the distribution will give a good indication of the value to apply.

Step 2 – Determining the histogram peak

The second step is the extraction of the peak value from the histogram. This is perhaps one of the most challenging areas, depending on the histogram form.

A ragged histogram with widely ranging values will give a poorly defined peak, we can’t just choose the highest value. We could produce a cumulative histogram and look for the mid-point defining the median value, or we could use the arithmetic mean. Depending on any asymmetry in the histogram we might get widely differing values. Issues can occur if the data is quantised. In the past we have used 14 bit data of a small satellite system with no quantisation and also 10 bit data from the same instrument which does show quantisation.

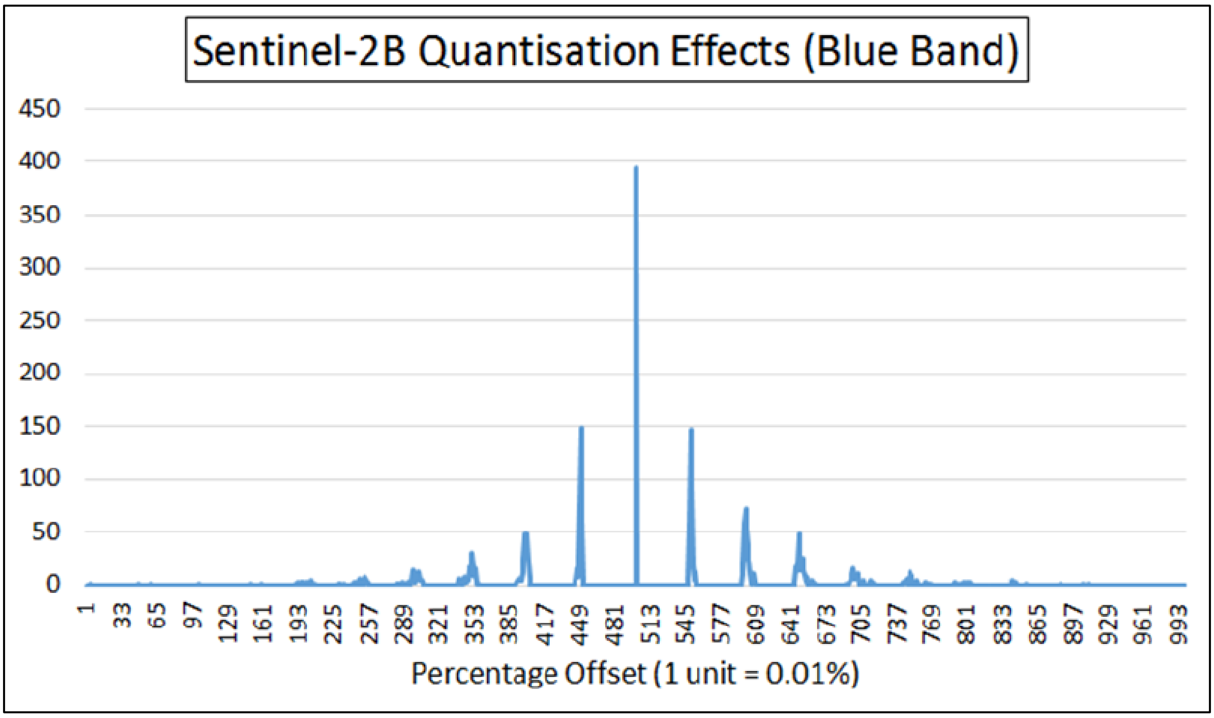


Figure 3: Histogram of pixel-pairs showing strong quantisation effects.

With quantised data the median value is more difficult to determine due to the step like structure of the cumulative curve given the gaps in the data values (figure 3), any shift of where the true peak position is placed will be seen as a calibration offset. In heavily quantised data there are two options. The central peak we cannot use, as it will say there is no offset. However, if we look at the secondary peaks to either side of the central peak they are generally asymmetric and consistently asymmetric from image to image for any image pair, we used this difference initially to prove to ourselves that consistent variations existed, although it is difficult to use these peaks to quantify the amount of calibration change required.

The reason it works is that for a perfectly calibrated system we get a central peak and an equal division of value to either side of this peak. If there is an offset, a small proportion of pixels will be pushed to the right or left, depending on the offset, producing asymmetric peaks.

Given we can’t use the peak heights directly we can actually consider a much simpler measure, if we ignore the central peak and use all the other data with an arithmetic mean, we found that the results were consistent and could be well-matched to the imagery. Therefore we believe that even with quantised data we can achieve reasonable results.

Step 3 – Eliminating Gradients

One issue we mentioned for single images, was that when we applied the algorithm, if we have a surface gradient in brightness, it shows up as a trend in the derived relative gain values. To eliminate gradients there are a limited number of choices. We could use an average of many images, which would re-inforce the smaller high frequency detector to detector variations we are interested in, while reducing lower frequency gradient induced features. This approach is fine and can be considered when a large amount of data is available. However, for monitoring changes with very high temporal resolution this approach is not feasible.

The alternative is to remove the low frequency resolution features and just highlight the high resolution features in the retrieved coefficients. Therefore we use a local polynomial based on the Savitsky-Golay method, usually five or seven point to remove the low frequency terms. The overall effect is that we can ignore the surface variability that is inducing the lower frequency changes and still recover consistent high frequency terms over different surface types as shown in figure 4.

Additionally the magnitude of these effects can be seen to match those over homogeneous surfaces. Confirming that most of the persistent residuals we derive are true surface effects and providing a direct means for correction.

2.1.4 Issues

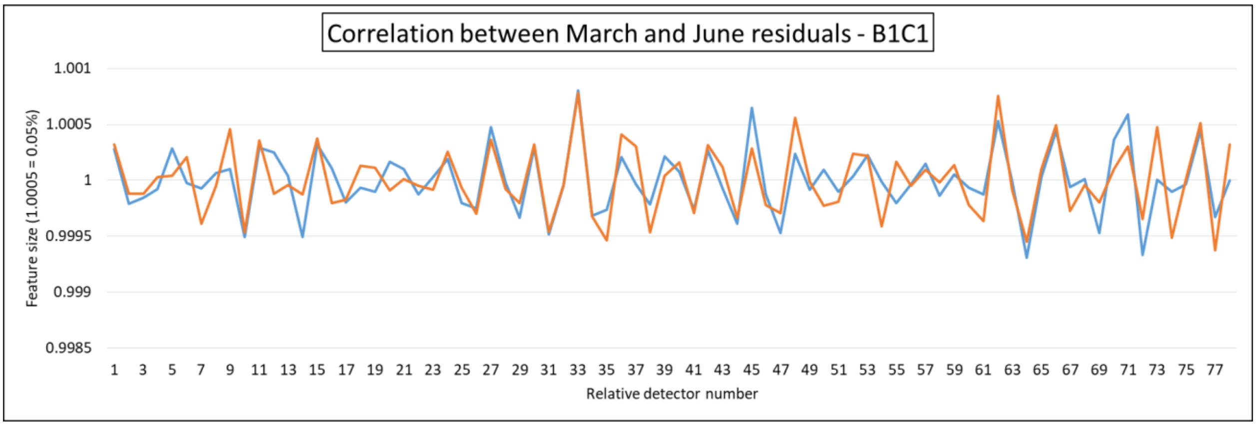
The biggest issues alluded two in the “Step by Step” section are quantisation and the variability of the surface in the imagery, especially the influence of clouds.

Quantisation we have discussed and an adequate approach based on ignoring the central peak data and using an arithmetic mean of the remaining data seems to provide a robust method for assessing the relative gain persistent residuals from one image to the next.

The second effect is more of an issue. The most important advantage of this methodology is the use of heterogeneous images in the analysis, providing very high temporal resolution. However, heterogeneous images by their nature make it much more difficult to extract the often very small relative gain differences. Large variations from one column to the next can produce “difficult” histograms that lack enough shape to be able to clearly identify the peak position and assess the relative gain. A problem often occurs with clouds. The algorithm is very effective with hard boundaries between objects, as although moving into a bright object causes a group of extreme ratio values, the impact they have on the peak is normally quite low.

However, clouds do not have sharp boundaries, the boundaries are diffuse, producing ramps in ratio values as the clouds are entered and exited that change the shape of the detector to detector ratio distributions and affecting the offsets. Obviously it depends on the amount of cloud, but also the contrast with the underlying surface. Some cloudy images distort the extracted values enough to significantly reduce the correlation between images.

The only way to mitigate these effects are to increase the number of images, this is especially true with the relatively low magnitude features we see in band 1 of OLCI. Examples of the size of residuals we can determine are shown in figure 4. The top plot is for band 1, Camera 1, the bottom plot for band 13, Camera 1.



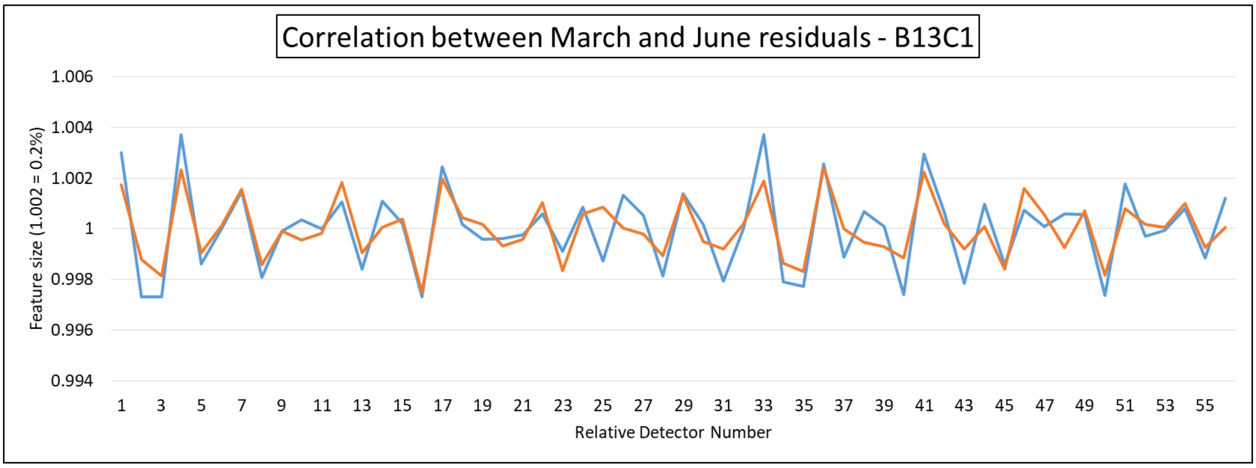


Figure 4: Two plots of the recurrent residuals from two different dates.

In figure 4, the blue line is the March 21st data, a whole day of data. The orange line is June 21st data. As can be seen, even as in the top plot for band 1, we have very small features (usually less than 0.05%) that we can see repeated from one date to the next. Individual plots show a lot of variability, but averages of a whole day show much more consistent response. This is ignoring the effects of non-linearity on the results (see non-linearity section) and any temporal changes over the three month period which may have reduced the correlation. The results suggest that these low level features and much larger features in band 13 (bottom plot, greater than 0.2% in some cases) are persistent over quite long time periods, but not detected during normal calibration using the on-board diffuser.

2.2 Non-linearity

For most imaging systems the non-linearity in response is a function of the detectors and the electronics (amplifiers, A/D convertors). Non-linearity is normally determined in the laboratory pre-launch. For Sentinel-2 for the VNIR bands a second order polynomial is fitted to reduce the effects of the non-linearity. Once in orbit the shape of the non-linearity curve is never changed and non-linearity effects are assumed fixed.

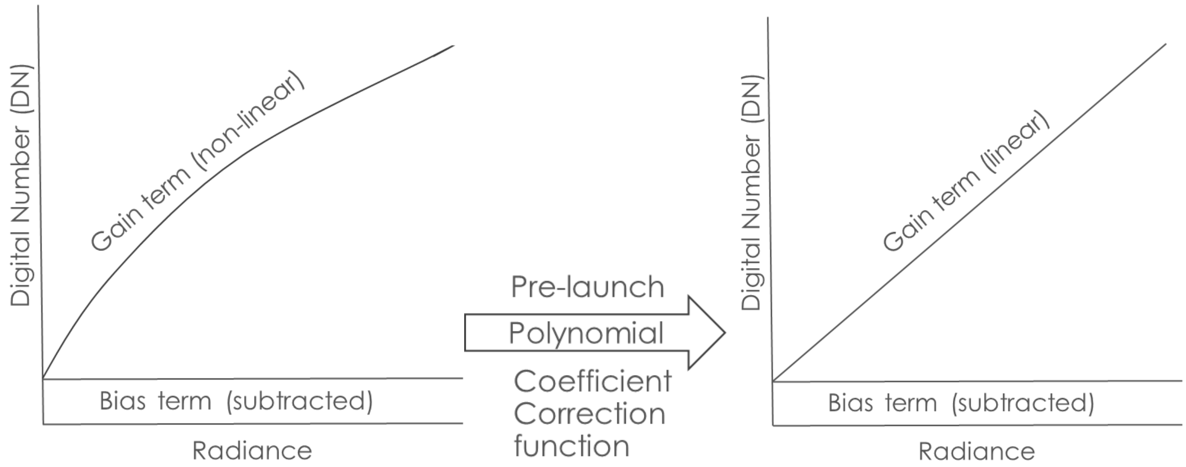


Figure 5: Non-linearity correction applies a polynomial to produce a linear relationship between radiance and the digital number.

In figure 5 (above) we see more or less how the non-linearity correction is performed. After a bias term is determined from dark images for each detector, it is subtracted from the detector DN response and then the remaining term will be a gain term relating the image brightness in digital numbers to the measured radiance in the lab for example.

Once launched from the evidence currently available, it seems that the non-linearity coefficients defined are never changed, the tie point of the diffuser for the upper part of the calibration is adjusted, but the non-linearity coefficients themselves are not. This leads to the possibility that the important part of the dynamic range of the instrument covering vegetation, soil and oceans is never observed fully and could have residual non-linearity effects.

This intermediate region (figure 6) is that where we are seeing persistent residuals from the relative gain analysis and it seems likely that there may be issues.

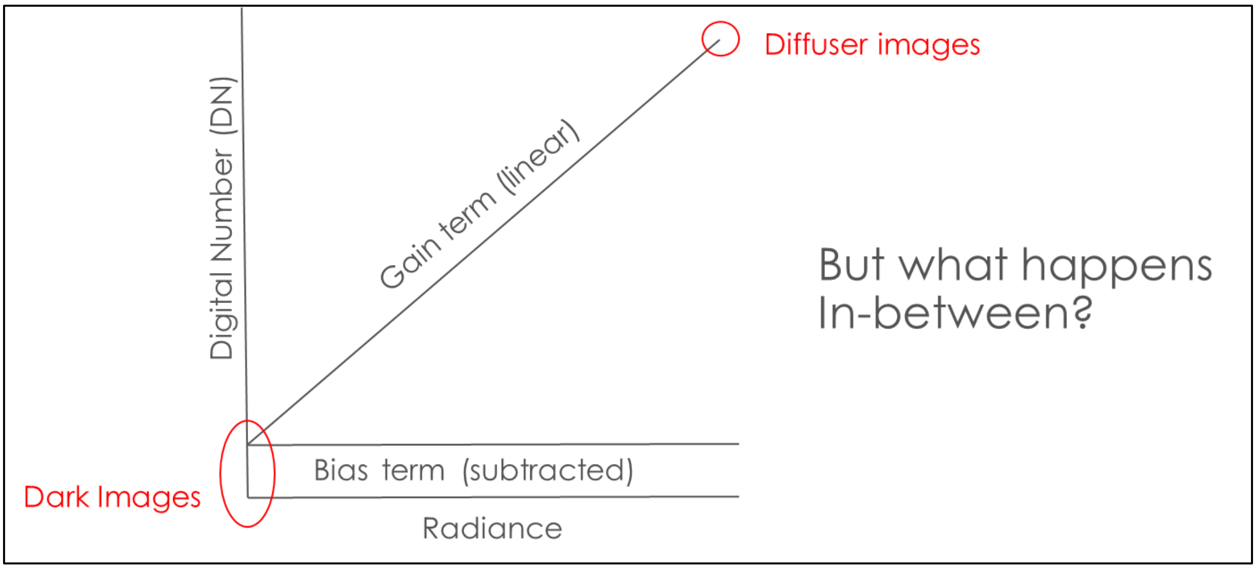


Figure 6: The intermediate region between the dark bias images and diffuser images is poorly monitored

If there are problems in this region related to non-linearity effects producing differences between neighbouring detectors, how would they appear?

The obvious sign is the effect of column striping that is evident for part of the column at a particular brightness then fades away and reappears as the image brightness changes. Obviously other causes such as a temporal effect in the electronics could introduce striping, but this would tend to affect all the detectors across the image array at the same time, while if we see localised effects for specific brightness targets it would indicate some behavioural differences between neighbouring detectors.

Obviously, if we can see the striping effects (normally only visible with features greater than about 0.2% of the magnitude of the signal), then we can detect these effects as calibration residuals, the magnitude of which is the average residual across the range of brightness's that make up the column.

So potentially depending on the non-linearity curves fitted for each detector, we could if there is residual non-linearity, determine it using an adaption of our relative gain algorithm, discussed previously.

2.2.1 Our Approach

This is an extension of the relative gain algorithm. It uses multiple **heterogeneous** images to build up a relationship between the persistent residual depth and the average column radiance for the target image for which the persistent residual has been determined. With enough images, we can start to clearly see a correspondence between the depth of any specific residual and the target brightness over the whole dynamic range of the sensor.

As with the relative gain algorithm we use Level 1 calibrated data as our input, the only additional parameter collected is the column average radiance or scaled radiance.

2.2.2 Principles

The principles followed initially are exactly the same as the relative gain algorithm. We determine the column to column ratios and define the ratio histogram, find its peak and eliminate gradients by focusing on the high frequency element. These are determined for each image or sub-image used. The granularity of our measurement is determined by the image size.

One thing we always capture when extracting the relative gain values, is the corresponding average column radiance for each column. These two parameters provide the basis for determining if there is a brightness relationship to differences between neighbouring detectors. In a perfectly calibrated system, no matter what the brightness of the target, neighbouring detectors will not show linear striping effects.

However, if for a certain brightness range (figure 7) there is a difference in behaviour between neighbouring detectors then the non-linearity can be observed as striping between detectors at specific brightness targets.

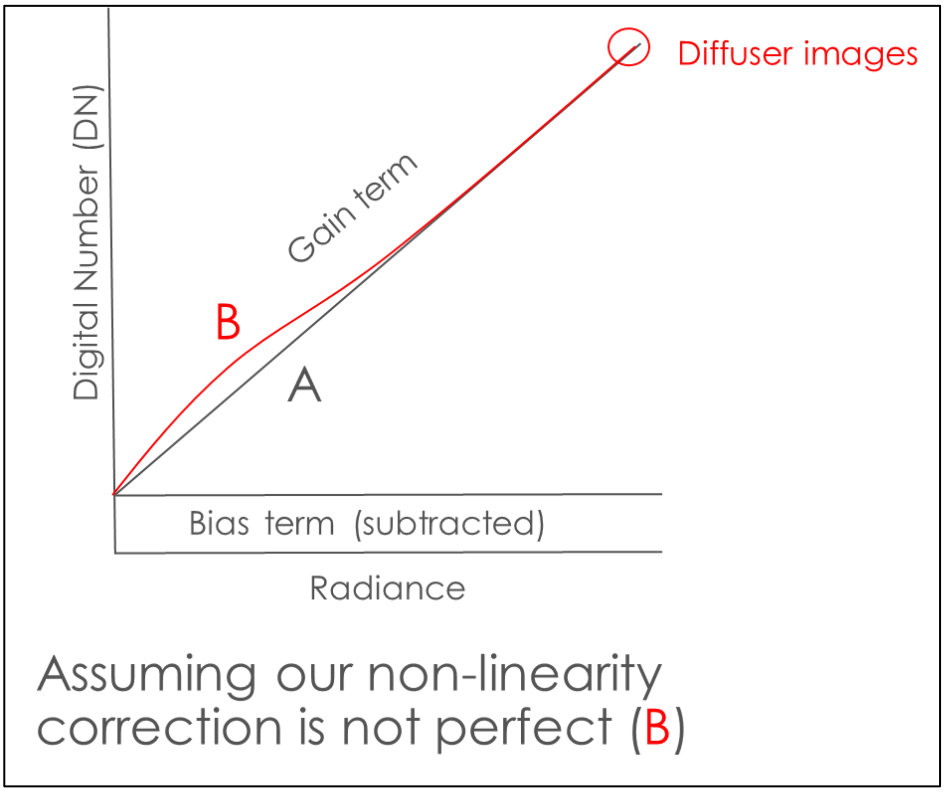


Figure 7: Comparison of a perfectly calibrated detector (A) to one showing non-linearity errors (B)

In figure 7 above, imagine we have several homogeneous images covering the whole dynamic range of radiance. As we ratio A/B we find that for really dark targets we see little difference, but as we move to brighter targets we see our ratio value decrease steadily to a minimum at about one quarter of our dynamic range. By the time we move to half the range the ratio values are more or less back to one and we can again see no mis-calibration between detectors.

In figure 8 we can see an example from Band 13 Camera 1 of OLCI, showing the intermittent nature of the striping produced by such an effect. Stripes over darker areas disappear into brighter areas. The stripes can be bright stripes or dark stripes based on the relative ratio.

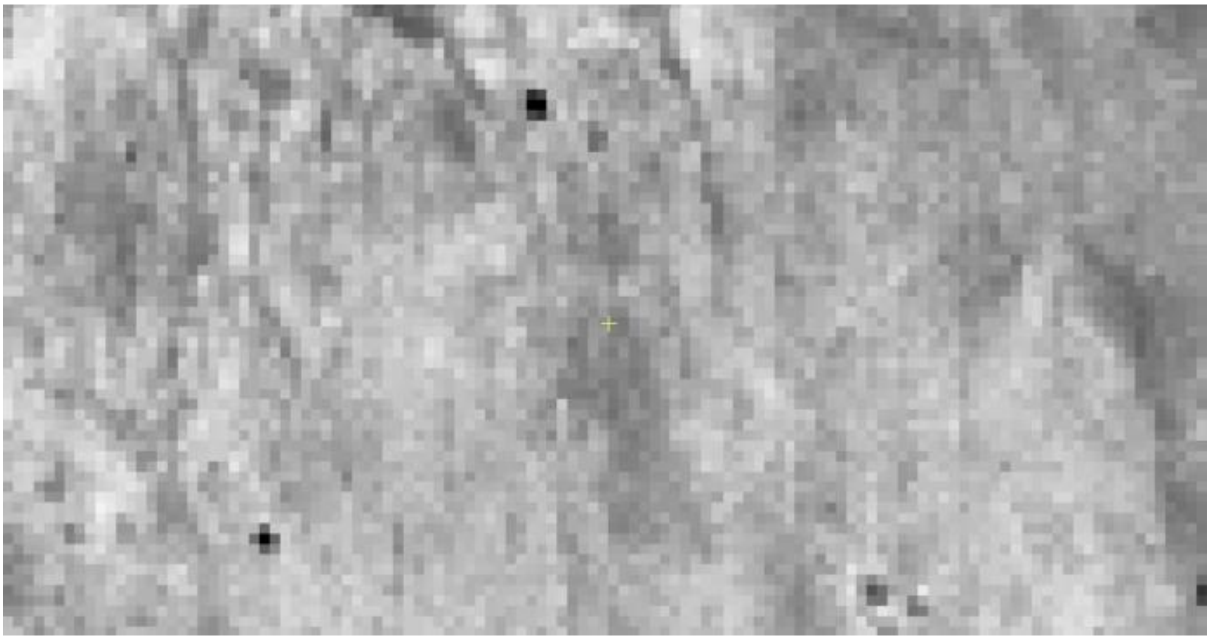


Figure 8: Band 13 Camera 1, notice the stripes are partial and fade in and out in the image

Using this model as a basis to predict behaviour, it is fairly clear with enough images we can plot a range of radiance values against the corresponding persistent residuals to determine if there are systematic variations in the residual depths with target radiance. There are however some difficulties, in that within a single image, for a single column pair we may find a range of radiance values and a range of residual values, which are averaged at the image level and this average is not necessarily a true representative of the relationship and will induce scatter in the final result. This is illustrated in figure 9.

In figure 9 we have three materials, A, B and C with radiance values of 50, 100 and 150 and corresponding feature depths of 0%, 0.4% and 0%. If we have a scene that is split between for example land with a value of 150 and sea with a value of 50, when we average the radiance we get 100, when we average the feature depth we get zero.

Now if we have a scene which is homogeneous made of B, we have the same average radiance value of 100, but a feature depth of 0.4%. So in effect, the varying relationships in brightness and feature depth when averaged to get a scene value, will show scatter in the final plot due to the inconsistencies described in the simple example shown in figure 9. So we need to consider how we deal with this issue.

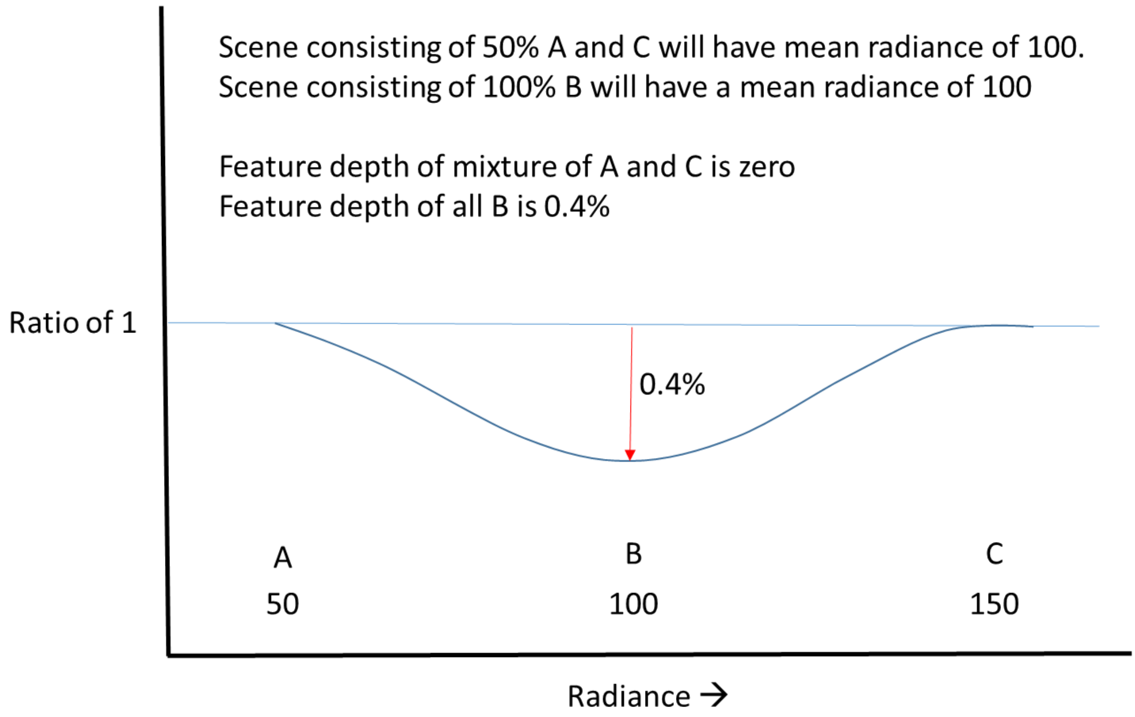


Figure 9: Showing that averaging feature depth and radiance does not necessarily give the right answer.

2.2.3 Step by Step

The initial three steps are the same for the relative gain analysis,

- Create a histogram for each detector to detector pair of the ratio values between neighbouring columns for the detectors to be compared.
- Find the peak of this histogram which is related to any relative shifts due to calibration differences between neighbouring detectors.
- Extract the high frequency component related to detector variations avoiding lower frequency gradients due to surface brightness effects.

The fourth step is related to plotting the high frequency values against brightness for each image or sub-image.

Step 4 – Generating the non-linearity plot

For each image processed for each detector we generate the relative gain value (high frequency) and the corresponding average radiance value for that detector for that image. We can then for each image, hence we need several images, plot the corresponding persistent residual feature against the average brightness value.

There are issues related to the quantity of images required to generate a suitable plot. For large well defined persistent residuals a small number of images would define the overall shape of the correction to tie any two detector responses together. However, for very small effects, a large number of images may be required and even these may be dominated by the scatter alluded to in the previous section.

However, it is feasible to try and extract the relationships. In figure 10, we plot an example from Band 13 from Camera 2 to illustrate the relationship for March 21st and June 21st.

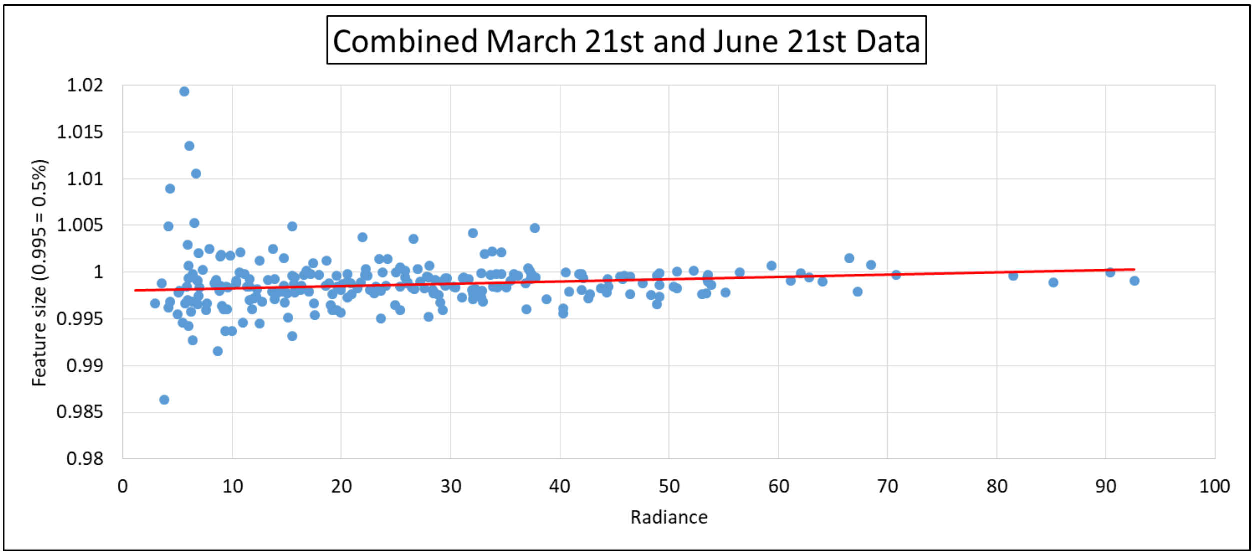


Figure 10: Combined plot of two days data showing a weak trend in for B13, C2, D222.

The example shown is not entirely convincing, the aim here is to merely illustrate that the effects we are looking for are very small and we need to consider the impact of the scatter on the final result. In figure 10 we can see there is no calibration residual for brighter targets, but it increases as we move to darker targets with features up to almost 0.2% close to zero signal.

Other examples show an almost continuous gain change, with a consistent value across the radiance range examined, with values exceeding 0.3% in Band 13 (figure 11).

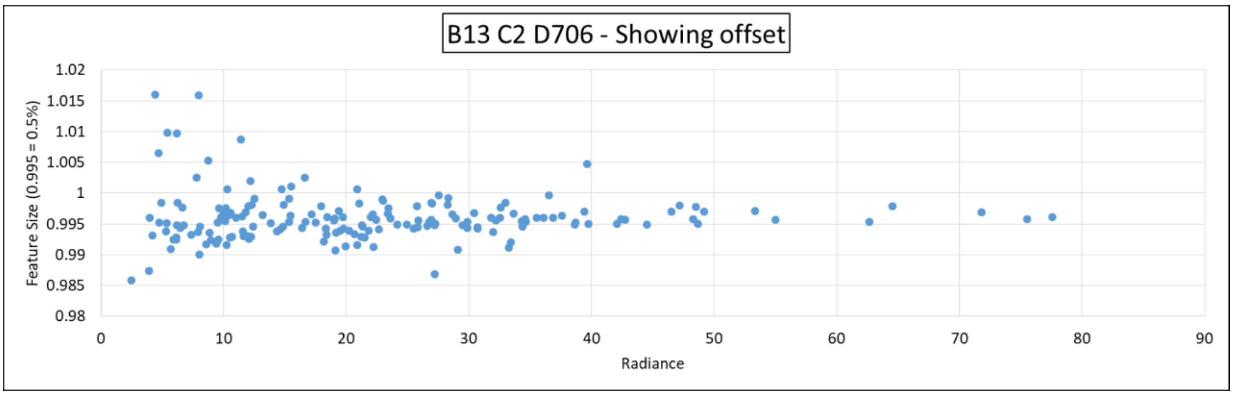


Figure 11: Persistent residuals across 180 images taken in March showing offset, average of 0.36%

2.2.4 Issues

The issues are related in part to the use of the relative gain algorithm, and the effects of using very heterogeneous images in the analysis. As mentioned in the section on relative gain, we can use large averages of residuals from a single day of data collection to produce good consistent values for the persistent residuals.

The second problem is how to deal with the averaging process over very different brightness targets with different magnitude persistent residuals as illustrated in figure 9.

The averaging process is required to generate the histogram, but the averaging process also is responsible in part for the scatter in the non-linearity plot. So one of the key elements to be examined will be how to extract the information we require to reduce these scatter effects. For the moment the approach we suggest to derive the first pass correction algorithms is to use several days data to get a good scatterplot to derive the relationship for the correction.

Once we have a relationship for every detector, then we need to derive the corrections and apply them, but this in itself is not as simple as it seems, as we do not know which is the best calibrated detector to use as a reference. In theory if we had one really well calibrated detector, we could say we are going to map every detector to this one and extend the relative correction out to all the other detectors from this absolute calibration reference. In reality we don't know which is best.

So we need to come up with a methodology that will reduce the impact. Our first suggestion is to use an area of the detector array which is showing little or no variation between detectors and use a detector in that area as the reference.

Then we need to migrate the correction out from this point, by either

(a) Combining correction factors
(b) Changing two values on either side of our reference, iterating and then moving to the next two.

The first option is easier from the point of view of computation. However, we are concerned that any residual error or precision effect may increase rapidly given the large number of detectors in each detector array.

The second option means re-running the relative gain correction several hundred times and changing one or two detectors at a time (one on either side of our reference). This would have to be repeated for all bands showing persistent residuals. So quite intensive processing.

2.3 Signal to Noise Ratio (SNR)

Regular measurements of the SNR can be important, not only to validate the pre-launch behaviour of an instrument once launched in space, but for the long term monitoring of the spacecraft health. On many systems with on-board calibration devices, SNR can be estimated on a regular basis (every few weeks in the case of OLCI on Sentinel-3), but many sensors do not have on-board systems, especially the new generation of small satellites being developed throughout the world. Additionally, if a major satellite system calibration device becomes degraded or fails, then the method described in this report provides a vicarious, alternative method to estimate SNR.

This data can not only be used to determine how well the spacecraft is performing, but also has the potential to be used in estimating the uncertainty on image products at various levels, as part of a comprehensive assessment, such as that proposed within QA4EO, where pixel level uncertainty on image or derived products is the ultimate aim. By characterising the variations in noise at the detector level in any image we have a parameter which can be used in the uncertainty estimates of any data product, as well as providing important quality information to ground controllers of large satellite constellations.

In this report the noise we deal with is assumed to be purely instrumental noise, so that the optics of the system are perfect and the only uncertainties arise from the properties of the detectors and electronics. It is assumed that the image values *x*, consist of a "true" signal, and an additive noise component. The noise is generated by the detectors and electronics on board, including dark current noise, read noise and Poisson distributed shot noise. The noise can be added during the read process of the detector or any later stage as the signal is passed through the electronics subsystems. The simple expression for the observed signal is:

$$x_i = t_i + \varepsilon_i(t_i) \qquad \text{Equ. 1}$$

where *t* denotes the true signal (that which arises from perfect imaging and optics), while the noise term ε is taken to be a random variable with zero mean, and variance:

$$E(\varepsilon_i^2) = \sigma_i^2 + k_i t_i \qquad \text{Equ. 2}$$

where the first term on the right hand side is independent of signal (but allowed to vary from detector to detector), and the second term denotes *shot noise*. In practice other sources of error in an image may lead to effects that appear as *striping* in an image. These are effectively multiplicative biases, and will be examined in a later report: the noise we deal with here is purely additive. The signal-to-noise ratio we are interested in is defined as:

$$SNR = \frac{t}{\sigma} \qquad \text{Equ. 3}$$

and will be itself a function of signal, given Equ. 2. This is the way the term is widely used in the remote sensing community; other definitions may be found in electrical engineering departments or among electronics groups.

Of course the optimal way to characterise the noise parameters in an image is to arrange for a uniform field of light to fall on the detectors; variations in the values read from the detectors then are all we need to examine. If we were to have such an image then noise characteristics could be found by examining the variability (essentially, the variance of the data values) and matching it to the expected form (Equ. 2). In some systems there might be an attempt made to effect such a situation, by the use of on-board calibration systems of some kind. Current satellite-based observation systems, designed to be small, light and to be launched regularly, cannot afford the expense, complexity and weight of such additional subsystems. Alternatives using the images that are actually taken from space, and an automatic procedure is preferred so that such analysis can be done autonomously – conceivably, on-board future missions.

For those satellite sensors that do not have diffusers, or other on-board calibration devices, the main difficulty in determining the SNR of a system is that we have to make do with inferences based on actual images of the earth: we must find a suitable way of determining the variation in a signal response without it being affected by variations in the image due to the surface or atmosphere. Many workers have therefore chosen "homogeneous" surfaces such as snow-fields or deserts to try and determine the noise contribution to a "fixed" signal background. In other words, we know the radiance of the surface and we assume no contribution from the surface or atmosphere to the noise. One of the major difficulties is that there are very few surfaces which are large enough that are truly homogeneous to allow an estimate of the SNR, especially for sensors with a large Instantaneous Field of View (IFOV) such as the UK-DMC2 sensors with a 640 km swath.

Another issue is related to the spatial resolution of the sensor, where a sensor with low spatial resolution may give an effective measure of SNR over a snow-field or desert site, while a higher resolution sensor gives an underestimate of the true SNR as it is sensitive to small surface variations (such as shadowing due to snow-dunes) that are not detectable and are averaged away in the lower resolution sensors. This limits the potential natural targets for such high resolution systems.

The academic literature on remote sensing is almost wholly concerned with extracting information from an image, or more general signal, minimising the effects of noise. No kudos attaches to the idea of studying noise in the presence of a confounding signal, and major journals generally avoid publishing such work. The exchange of ideas and methodologies between those for whom calibration is the principal concern therefore tends to form a very "grey" literature. Specialist conference proceedings, reports to Agencies, and other difficult-to-obtain sources are the main means of dissemination of this kind of work. A full-blown literature search is therefore not possible, so this element of the background must be based largely on personal experience and informal communications.

All the methods applied based on literature search tend to use similar sets of steps:

Use of a small moving window to gather local statistics, and identify areas showing low levels of variability (most homogeneous surfaces). The mean values and standard deviations over those areas are then extracted and used in the determination of the SNR. In some methods the technique of area growing is deployed to generate enough pixels in an area to use in an assessment. An interesting variation on the standard approach was proposed by Lee and Hoppel (1989) namely a method using small moving windows in which the mean value (squared) and the variance were plotted against each other and a straight line drawn through the data points, using Hough transforms to estimate the line position and from this the derived SNR. This assumes no dependence of the noise on signal. This method has been tested and in some circumstances works very well, but the results from our tests were very inconsistent from image to image and highly dependent on the image content.

Most methods by their nature have limitations. They tend to choose the most homogeneous areas and, in consequence, overestimate the SNR through selection bias; they generate only spot estimates across the FOV, and thus not a complete knowledge of the SNR; they cannot give detector by detector knowledge from each scene processed.

The selection bias arises from the fact that the calculated variance over a number of pixels for which the signal level is identical will itself vary from one group of such pixels to another. The expected value of such a variance will be the variance of the underlying noise, but there will be some sort of variation and choosing sets of pixels with the smallest observed variance will underestimate the true level of image noise.

2.3.1 Our Approach

EOSense has developed a concept that uses a moving window to extract scene statistics. Normally the SNR of an earth observation sensor is determined by observing a homogeneous surface and taking the mean and standard deviation of the surface to get a signal value and corresponding noise standard deviation.

This is a single data point. We would normally use multiple data points to define a relationship between the radiance measured and the corresponding noise standard deviation, with the noise decreasing as the signal decreases.

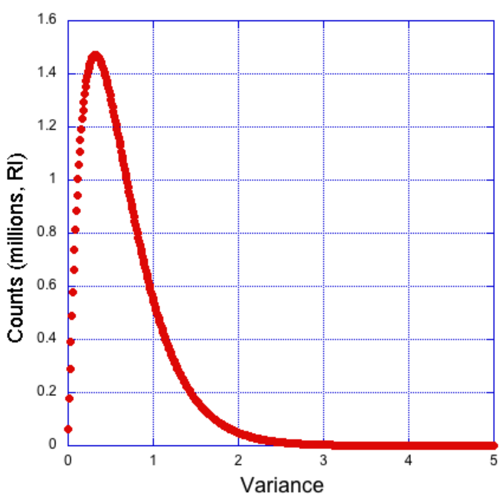
For practical purposes it is very difficult to find a range of targets that meet the requirements to generate these points, especially during the commissioning period for new satellite systems. As an alternative we have developed a new method of assessing the SNR of any satellite EO system by collecting statistical information of heterogeneous images to try and determine the effective SNR profile.

This is not just a single point measurement; this approach tries to recreate the true radiance to noise relationship across all radiances observed by the system by using multiple images that can run into the hundreds. In fact the more images used the more reliable the result.

2.3.2 Principles

We have taken two main approaches to the problem of characterising instrumental noise in satellite image. The first avenue is similar to others we have seen discussed above. Here the variance of a small number of contiguous pixels is calculated, repeatedly across the image. This is done in the belief that for at least some of these samples the variances will be dominated by the variance of the instrumental noise, with scene variability contributing little or nothing. The values are collected up into narrow bins, the contents of a bin at variance value *v* (say) being the number of times a variance in the range ($v-\delta v, v+\delta v$) is encountered. From the resulting histogram we can estimate the underlying noise variance in one of two ways: by Maximum Likelihood estimate, or by matching the low-*v* portion of the histogram to values expected theoretically from pure noise. The ML calculation also gives us a (rather too-low) estimate of the uncertainty of that calculation through standard statistical theory (Cramer-Rao). The curve matching also gives us an estimate of the error, which we think may be a novel introduction. The one thing we do not do is to form any kind of average of the *v*-values in the low-*v* part of the histogram, as this introduces a serious selection bias.

The approaches just described work best if there is a reasonable supposition that, over at least some parts of the image, the true signals for adjacent pixels are effectively the same: that there are uniform patches of the land surface within the swath of the instrument. The plausibility may be shown by considering the histogram of *v*-values from a high resolution panchromatic image of part of the Libyan Desert. The IFOV corresponds to a ground element of about 1 meter, there is a single cover type present, and conditions would seem to suit such an approach. The histogram is shown in figure 12.



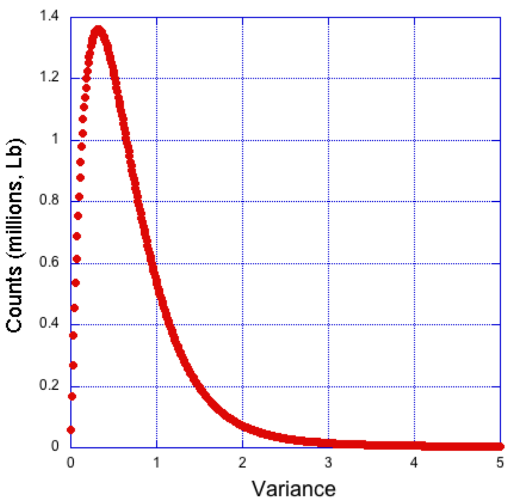


Figure 12: Distribution of variances from a random image (left) 5 pixels per sample. Corresponding histogram of *v*-values from an image of Libya4 (right)

The similarity of the two curves is striking, and the correspondence is strongly brought out when we examine the scatterplot of the two sets of values (figure 13 left, for just the 30 lowest bin values) corresponding to those portions of the plots to the left of the peak. Figure 13 (right) shows the result of scaling the theoretical curve (blue) to the observed set of *v*-values (red).

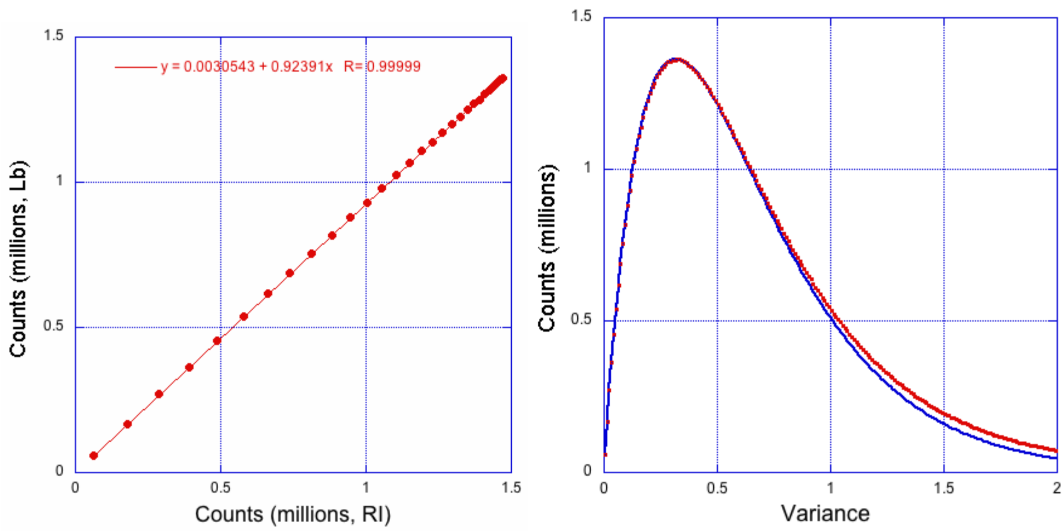


Figure 13: Scatterplot of random and real data (left) and overlaying scaled theoretical data on real data (right).

An alternative approach is a technique from geostatistics that enjoyed some popularity a generation ago among part of the remote sensing community, usually as a means of estimating scale (correlation length) over the land surface. This is the use of semi-variances, that is the mean squared difference in signal of two points a given distance (lag) apart. At large lags this value is more or less constant, but decreases with decreasing lag. Interpolating the semi-variances over small lags to a hypothetical lag of zero does not give zero, but instead what the geostatisticians refer to as "nugget variance", which we would interpret as detector noise. Our first studies on this approach were not promising, as the classically-calculated semi-variances showed too much scatter at small lags. It is easy to show that this method can be affected by individual bright or dark pixels. We think a further problem is in calculating the numbers directly, giving too much weight to large differences, when more consistent results are emerging when we use a histogram matching approach to estimating the variance at each lag. We believe we can justify the methodology in terms of a simple model of surface variability we have developed, and which we have not encountered in any previously published work (though our experience of geostatistics is limited).

An alternative model is that the surface-leaving radiance varies only slowly (between obvious boundaries), whereas the noise is uncorrelated form one pixel to the next. Separating the image into a smooth and a very unsmooth components may then be a feasible means of estimating image noise. It's possible that Fourier methods could be applied here, assigning high frequency components to noise and low frequencies to "true" signal, but this introduces the problem of deciding on a threshold frequency, or an apodising function. Instead we have looked at a method which finds the smooth signal as that which best matches the observed signal, while satisfying certain smoothness constraints, and such that the residuals (noise terms) are uncorrelated from one pixel to the next.

This is essentially a penalised least-squares method such as used for image restoration in a number of disciplines. There is usually a free parameter (arising as a Lagrangian constant) in such schemes, this we can tie down by the requirement that the residuals be uncorrelated. The method has been developed only for one-dimensional data sets, which is fine as we can apply it to the long series of data from a single detector. It is not clear how easily this could be developed into a two dimensional algorithm, or what the computational cost would be. We have made a limited number of tests with this algorithm, and get results compatible with those from our other approaches, although it is evident that the least squares formulation again leaves us vulnerable to outlier pixel values. Some form of image pre-processing may need to form part of a fully workable method.

Our fifth and latest scheme is not to work with variances at all, even though that is what we are attempting to recover. If we take not the variance of a small set of pixels, but instead a linear combination of them, chosen to nullify a constant signal, then the result will be a function of the image noise only. The distribution of the results can then be fitted against the expected distribution; in the case of the high-resolution images we have been working with, this is a simple Gaussian. This approach has the main advantages of the histogram based methods – automatic elimination of samples containing boundaries, or over very heterogeneous areas – as these fall naturally to extreme values, which are ignored.

2.3.3 Step by Step

For Histogram methods, a "configuration" is decided upon: this could be a window of size 2x2 pixels, say, or a set of pixels in a row, or 3 consecutive pixels along a column (and thus from the same detector). A bin size is selected for the histogram: this should be quite narrow, as it can be coarsened if needed by combining bins, but it cannot be refined. Three arrays are set up of length MAXBUF (say), these we label COUNTS, MEANS, and VARS, all initialised to zero. If the bin size is h then the maximum variance we play with is Vmax=h*MAXBUF. We pass through the image, and for each set of pixels we calculate the variance of their signals, and also their mean value. If the variance is less than Vmax then we increment the count of the appropriate array element by one, add the mean value to the number in the corresponding element of MEANS, and the variance value itself to the corresponding element of VARS. At the end of this we divide each element of MEANS and VARS by the corresponding element of COUNTS. The output histogram is essentially a table with MAXBUF rows. The first column of the table contains the midpoints of the histogram bin, the second contains the endpoint of that bin. The next is the contents of COUNTS, and then the modified numbers from MEANS and VARS. Generally it is found that the first and last columns are almost identical, but for the first few bins, those with the lowest v-values, there may be observable differences. This is an indication that the histogram could be affected in those values by quantisation problems.

For the ML method we choose a bin, and calculate the mean value of the counts and the mean value of the vars variable for all bins up to the one chosen. These are readily reconstructed from the table. The ML estimate is now the solution of a transcendental equation with these two numbers as parameters. It is straightforward to calculate ML estimates for any portion of the histogram, to help check for consistency. The mean signal of the bins used – again, readily calculated from the array table – is the appropriate mean value to combine with the noise variable to estimate SNR.

For the histogram matching our aim is to compare the numbers in a portion of the COUNTS array to members of a one-parameter family of curves, depending on the underlying noise distribution. For Gaussian noise these are χ^2 curves. After much experimenting we believe that the most sensible approach is to scale the counts in our selection so that they sum to one, and to do the same with the scaled integrals of the χ^2 function over each bin. The squared differences of these two sets of numbers, when added, form an objective function; the desired noise variance is that which minimises this. An uncertainty estimate can be determined for the result.

The geostatistics approach we deploy is not quite standard. Semi-variances are classically calculated by summing the squares of differences f signals, but as mentioned above this can be badly affected by bright or dark isolated values. Instead, a histogram of the differences at each lag is generated, and the lower values in it are fitted to a Gaussian (in practice, we fit the log of the counts to a quadratic). This is showing improved results over the earlier work, although testing has only recently begun on this.

The same methodology is adopted in our most recent scheme. We pass through the image, calculating certain linear combinations of sets of successive pixel values, and accumulate the results (which may now be positive or negative) in a histogram; this is then fitted to a Gaussian and the variance of the Gaussian is the noise we infer. Unsuitable numbers never get included in the calculations as their derived numbers are too high.

2.3.4 Issues

The histogram matching technique has been extensively tested on images of the Libya Desert for a particular instrument. The results have enabled us to confirm a linear relationship between radiance and noise variance, and are generally consistent.

The most significant problem we have seen is that if the data is highly quantised (by which we mean the digital numbers are squeezed into a smaller number of bits, before conversion to radiance values) then the histograms can become very distorted if the number of pixels per sample is small. This makes the whole fitting process difficult. Having said that, the last method we discussed seems more robust than the others examined and this will be tested on highly quantised data for Sentinel-2 in the near future. For Sentinel-3 OLCI data the quantisation was not apparent in the Level 1 data, so we may be able to try several different algorithms to assess the noise and give comparative figures.

The second issue we need to explore is the problem of spatial resolution. Most tests have been performed on very high resolution systems. However, we do have enough evidence from medium and low resolution systems to give us confidence that we can extract some useful information (previous limited tests on AATSR and MERIS). We will get a clearer view on whether this is a major issue as the project develops.

Finally, the aim for all our work has been to use heterogeneous images, normal images, in our analysis. The SNR histogram methods show that when we have very heterogeneous images, the histogram shape becomes distorted due to the effects of mixing different surface components of very different brightness, producing large variance values with poorly defined peak values. This may limit the usefulness of this method to sites which show only moderate variations in surface brightness. This will be examined as the project develops.

3. Part 2 – Level 0 Processing

- Drift of on-board calibrator
- Sensor drift monitored relative to the earth without and on-board calibrator

These two algorithms are variants of each other. They are very different from the level 1 algorithms discussed in Part 1 of this document, in that instead of using a very limited number of images to determine parameters such as relative gain, non-linearity and SNR they instead use a very large number of images (in fact as many as possible) to try and assess changes in sensor performance.

Global datasets are preferred, so there are no regional biases to the sampling of the data. Both algorithms are based on creating a new calibration reference, by which we can see either changes to the on-board calibration devices or direct changes in the sensor performance.

No absolute calibration drift assessment can exist without some form of reference calibration target. In our case this target is the whole earth, with the underlying assumption that the spectral albedo of the bands of interest of our sensor is not changing during the lifetime of the sensor and hence we can adequately track changes in either on-board calibrator or the sensor itself by direct reference to the level of signal change (after bias subtraction) of the level 0 data.

3.1 On-board Calibrator Drift

One of the issues with any on-board calibration system is the tendency for drift to take place in the response of the on-board device. Landsat 8 OLI has multiple calibration systems, these provide a good estimate of the calibration stability, but still show small, but different drift behaviours (Markham et al., 2014). An alternative approach is to use vicarious calibration sites on the earth's surface as targets from which we can derive the drift of a sensor (Mishra et. al, 2014). These use Pseudo-Invariant Calibration Sites (PICS), areas of the earth that have shown good long term stability over a 30 year time period. They key element is the stability of these sites, there are however problems in use including extensive modelling of the surface BRDF, atmospheric variation including the presence of clouds over the sites and the relative paucity of data given instrument revisit times. An example of the complexity, is the processing used for ATSR-2 and AATSR for the on-board VISCAL device, the same device used in SLSTR.

As part of the validation process of the (A)ATSR on-board calibrator the Rutherford Appleton Lab (RAL) carried out a calibration exercise of the North African Pseudo-Invariant Calibration Sites (PICS) and snow sites in Antarctica and Greenland using Level 1 Radiometrically Corrected data (Smith and Cox, 2013). In this case a reference reflectance model based on ATSR-2 over these sites was used as a basis for comparison, which does not require time-coincident data. The data is restricted to nadir view only and the BRDF can then be considered as a function of solar zenith angle and the measurements collected can be fitted using a polynomial.

Once ingested the images are checked for location, cloud screening with checks on spatial uniformity over the targets using a windowing method based on thresholds. For accepted images they compute,

- Mean reflectance
- Standard deviation
- Minimum reflectance
- Maximum reflectance

- Mean solar zenith and azimuth angles
- Mean view zenith and azimuth angles
- Total number of pixels in the scene (clear and cloudy)
- Number of cloud free and non-saturated pixels in the scene used in the average.

The next stage in the RAL processing was to account for any long term drift, as the ratio between the measured reflectance and the reference BRF, this was carried out for the PICS desert sites, Dome-C in Antarctica and a site in Greenland, the same drift was observed in all sites which strongly suggested it was a calibration drift rather than a site specific variation.

In detail, the average drift in a specified time window, within that time window any observations that are more than two standard deviations from the mean were excluded and the mean recalculated, several iterations were performed until a stable result was achieved, usually within five iterations. The time window to use for averaging was determined by visual inspection of the results which removed most of the high frequency noise but without aliasing the low frequency drift. For the analysis the time window was 120 days, this could be reduced but at the cost of reducing the number of measurements in the average, as some sites had no cloud free observations in a given month. (Smith and Cox, 2013).

The drift values were saved to a lookup table and made available to the community. As can be seen the standard PICS approach is very time consuming with many steps, models and corrections including removal of outliers before an effective assessment could be made.

3.1.1 Our Approach

The approach developed by EOSense and tested on AATSR is based around using an alternative reference to the pseudo-invariant site, replacing it with a pseudo-invariant Earth. This approach, particularly suited to global data sets is based on the idea of using the whole earth as the reference data set. The obvious underlying assumption being the need for stability of the spectral albedo for the spectral bands we wish to monitor over extended time periods. There are several factors that can affect the "invariance" of the earth as a reference.

- There will be seasonal variations in the global albedo, so we cannot directly compare one month's data directly with the following month. Secondly, from year to year, depending on seasonal variability, we will have variations in any particular time period due to extent of snow cover, unseasonably warm or cold weather and cloud distribution. However on an annual basis we expect consistent average values.
- There is a variation in solar irradiance and earth-sun distance. The sun-earth distance varies by 3.3% during its orbit around the sun, producing a well-defined variation in solar irradiance at the earth's orbit which can be corrected for if we know the day an observation was made. Additionally the actual output from the sun, solar irradiance ignoring changes in earth sun distance, has been the focus of many papers and discussions. The magnitudes of these effects are very small, with estimated trends between -0.008% per decade to +0.037% per decade (Willson, 2014).
- The atmospheric conditions are certainly not constant from day to day over the same target on the globe. The angle of illumination and view may change to each target type and hence induce Bidirectional Reflectance effects. However, these can be considered high frequency "noise" effects when using a global data set over several years, as we will cover all atmospheric conditions and all illumination conditions.

However, we can estimate the uncertainties induced by these variations and in many cases it is possible to remove or reduce some of these effects by using multiple bands and multiple sensors. If this is the case the use of the earth as a reference may provide a simple method for assessing the long term radiometric stability of optical sensors.

3.1.2 Principles

In terms of methodology, we use two simple examples, a system with an on-board calibration device (such as a diffuser) and a system which has no on-board calibration device (this will be discussed in the sensor drift section later in this document).

In the case of the first example with an on-board diffuser we are effectively comparing two reference targets, the diffuser and the Earth. If the diffuser is unchanging and the Earth is unchanging on average over a year, then a simple ratio of the diffuser average and the Earth average value, after bias subtraction, will give a constant. As long as there is no drift in either reference, the ratio value will remain unchanged. Multiplicative changes in the detectors, electronics and optics, or changes in the digital gain will have the same impact on both the data from the on-board calibration device and the imaging data collected of the earth, hence will have a constant ratio value. If however, there is a change in either the diffuser or the Earth (or both) we will see a change in the ratio value proportional to the drift.

3.1.3 Step by Step

The processing steps will be as follows,

1. The Level 0 data needs to be extracted using a purpose written program. The average value for each column for both the forward and oblique views of SLSTR and nadir view for all five cameras for OLCI will be extracted for each orbit for each band and the bias value (blackbody for SLSTR and shutter for OLCI) subtracted from each value.

2. The corresponding VISCAL values during captures over the South Pole need also to be averaged and lower blackbody value subtracted from each average during the VISCAL collection for SLSTR, while the current calibration bias values need to be subtracted from the OLCI data.

3. Initially we propose, for the data values for each band (plus forward and nadir views of SLSTR), all the column values are averaged together to give a single orbit value, then all the orbit values for one month are averaged to give a monthly value.

4. For the SLSTR VISCAL values all the orbit values for one month are averaged to give a monthly value. For OLCI the regular diffuser average results are used in a monthly average.

5. In essence it is simply averaging the RAW counts from the data over a whole year and the RAW counts from the on-board calibrator over a year. If we now ratio the two sets of counts, if nothing has changed in the calibrator, we will see a constant ratio, if something changes in the detectors or main optical chain then it will affect the calibrator and Image data. If something changes in the calibrator alone, then there will be a divergence between the calibrator averages and the image data averages.

A first trial of the algorithm was performed on AATSR data in a previous study, using the first year of data and VISCAL (2003) as a starting point, we determined the percentage drift by simply calculating the ratio of the values against the first value in our set. We validated the results of the analysis, against the established measurement carried out in a different manner using Level 1 data. In this case, we referenced our results to those obtained by the Rutherford Appleton Laboratory (RAL).

The results are shown in figure 14. The dotted lines are the results from RAL based on the PICS analysis and Dome-C and Greenland data analysis, while solid lines are from the Earth Reference Method are based on divergences of the ratio between the average annual measurements of the spectral albedo for each band with the average annual VISCAL response for each band, both after bias subtraction.

The two sets of data using very different approaches in their processing show remarkably consistent results. This tends to validate that the approach developed and used by RAL gave a very reasonable model of the drift of the AATSR on-board calibrator. However, there are differences, notably in the 1.6 μm band although of the order of only 1%.

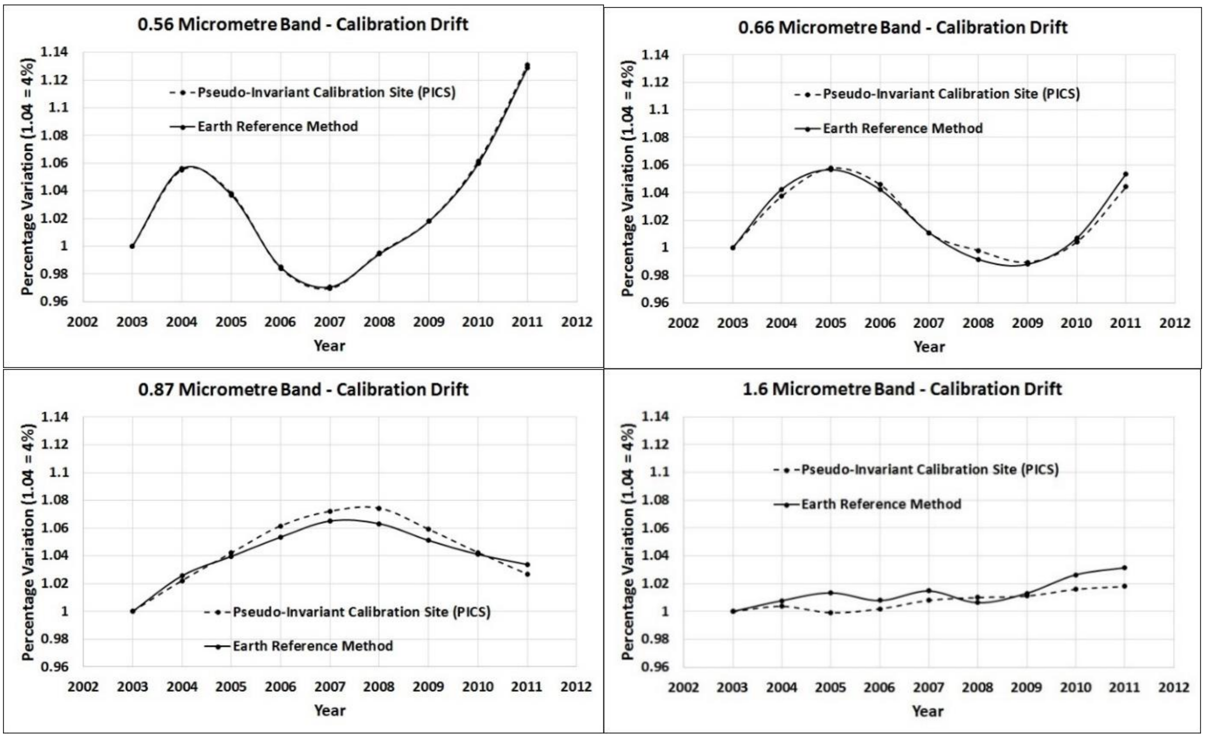


Figure 14: Final results comparing the calibration drift table (dotted) against the earth reference method (solid)

The advantage of the Earth Reference method is the simplicity of the approach, there is no data pre-selection, no conversions, no site selection, and no correction for BRDF or site conditions (cloud, snow, and atmosphere). It is a simple averaging of the RAW data counts after bias subtraction and provides a viable validation method for the established methods.

The overall uncertainties we expect not exceed 1% based on the instrumental design which is a similar order of magnitude as that from established methods using vicarious calibration over the Pseudo-Invariant Calibration Sites (Smith and Cox, 2013).

3.1.4 Issues

One of the two major issues related to the methodology is the problem of extracting monthly values against the varying signal background. Annual averages should be consistent, but we need to understand how the relationship varies from month to month to make adequate predictions on a monthly basis and perhaps even week to week.

The second problem is that we are assuming linear behaviour over the whole dynamic range of the instrument. Any non-linearity with signal level is not accounted for, as there is no pre-processing. However, we need to consider that for an annual measurement, assuming the signal levels are changing slowly, the overall mean values will be about the same and any effects due to non-linearity will be limited. It may be worth considering the current non-linearity correction applied and assess by simulation what that might mean to the calculated signal levels.

3.2 Sensor Drift Monitored Relative to the Earth Without an On-board Calibrator

In this case we are purely monitoring the sensor drift with time. If we use an example of AATSR (figure 15) we can see that over a relatively short period of time the signal level recorded dropped dramatically, with a gain change of 25% of the signal level being applied to recover the loss in response.

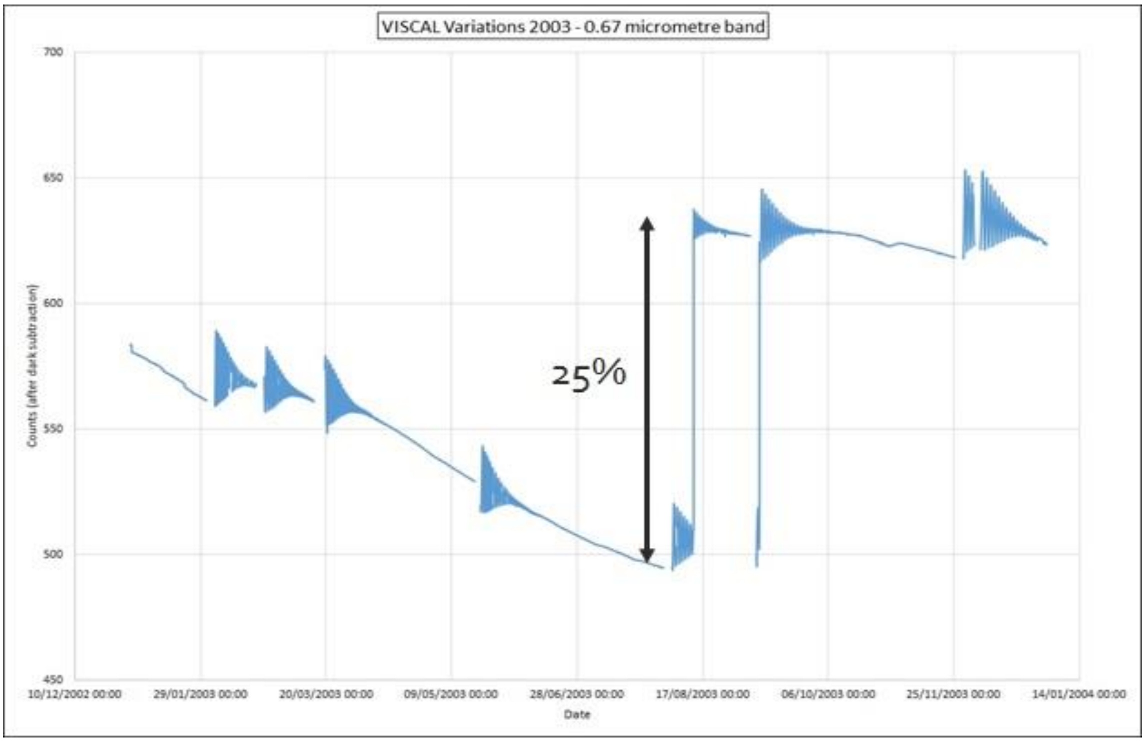


Figure 15: Large change in signal level during the first year of life of the sensor

The standard way of monitoring the loss of response when there is no on-board calibrator is to use a vicarious method, a cross-calibration approach over PICS using a well-calibrated reference sensor. Using a methodology similar to that used by RAL in assessing the calibrator drift in the previous section.

3.2.1 Our Approach

EOSense uses an extension of the method described in the previous section for monitoring the on-board calibrator, except in this case, we do not use a ratio of the earth reference data to the on-board calibration data, but instead use the earth reference data alone. Imagine we have a perfectly calibrated sensor that is unchanging with time and the earth reference on average is also not changing with time. In theory we will have a repeating signal, capturing the normal seasonal variations, every year the same magnitude, an example is shown in figure x for MERIS data red band.

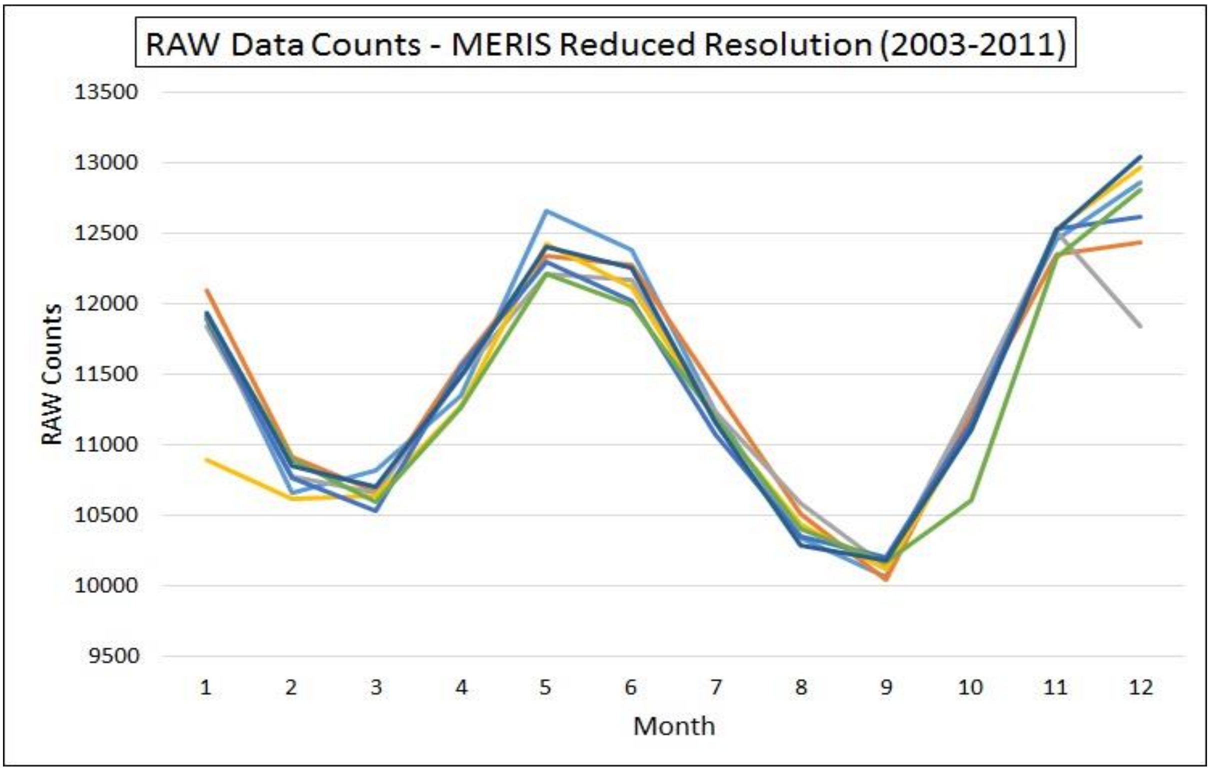


Figure 16: The repeated pattern of raw data counts during the annual cycle

As can be seen in figure 16, the pattern is repeated each year (different coloured lines are different years), the most variability we see is in December, possibly related to snow cover variations. But on average it is remarkably consistent. Careful analysis shows that the MERIS curve did gently drift downwards over several years then stabilise.

So in theory just by using the bias subtracted counts we can get a good indication of drift of the sensor.

3.2.2 Principles

For the second example where no calibration device is present, we can still measure the degradation of the system and corresponding loss of signal, by looking at the average earth signal and how it changes over the lifetime of the sensor, assuming of course that the earth's albedo is not changing. However to measure this change we need a full annual curve which does not show degradation as a reference to begin with.

In principle, given that the sensor is constantly changing, this is not feasible. However, we can in the case of OLCI get a very good approximation if we use the primary and secondary diffusers to correct the first year's data and apply those changes to the bias subtracted data to generate the curve. Once we have the initial curve we can then assess the changes in the calibration in the second year, which we can validate using the on-board calibration devices. We can then create a second year corrected data set which may have small month to month variations. By collecting multiple years, we can sum the corrected curves to remove small month to month variations and improving our reference curve.

This will work very well for OLCI. However for SLSTR given the large diffuser drift observed for AATSR we need to consider an alternative. In this case we can use the OLCI data as a proxy source of information on the reference curve shape, choosing OLCI bands that align with the SLSTR reflectance channels and generating the reference curve from OLCI data to derive the form of the curve and then superimposing the SLSTR results and adjusting accordingly to derive the sensor drift.

3.2.3 Step by Step

The processing steps will be as follows,

1. Generate a first year reference curve from the OLCI data for each spectral band. This will use the secondary diffuser on-board to provide a drift free reference for each spectral band. For SLSTR, OLCI bands will provide the reference shape in the first year, the reference magnitude determined from the SLSTR data itself.

2. The Level 0 data needs to be extracted using a purpose written program. The average value for each column for both the forward and nadir views of SLSTR and nadir view for all five cameras for OLCI will be extracted for each orbit for each band and the bias value (blackbody for SLSTR and shutter for OLCI) subtracted from each value.

3. Initially we propose, for the data values for each band (plus forward and nadir views of SLSTR), all the column values are averaged together to give a single orbit value, then all the orbit values for one month are averaged to give a monthly value.

4. The drift for the first year can be estimated by referencing back to the uncorrected first year data, which should recover the same drift estimates as derived from the OLCI on-board diffusers.

5. For years following, we use the initial reference curve to calculate the month my month corrections required and compare those results against the secondary diffuser data from OLCI to validate that the results are consistent and estimate the uncertainties of the comparison. The SLSTR data will be compared against its reference data curve, although validation may be more complex.

6. After the cross-comparison the corrected second year data will then be merged with the reference curves, reducing the effects (eventually) of month to month seasonal variations.

3.2.4 Issues

There are potential issues related to using the OLCI data as a proxy for the generation of the SLSTR reference curve. Also as mentioned in the previous section we are ignoring non-linearity effects, which could potentially become significant if the sensor response changes greatly during the first year of life to a point that the effective response of our detectors/electronics is different due to non-linearity effects. Simulation based on knowledge of the non-linearity corrections currently employed may help resolve this potential issue.

4. References

Lee, J.S. and K.Hoppel. "Noise Modeling and Estimation of Remotely Sensed Images" [sic], in: *Proceedings of 12th Canadian Symposium on Remote Sensing* 10-14 July 1989. DOI: 10.1109/IGARSS.1989.579061

Markham, B.; Barsi, J.; Kvaran, G.; Ong, L.; Kaita, E.; Biggar, S.; Czapla-Myers, J.; Mishra, N.; Helder, D. Landsat-8 operational land imager radiometric calibration and stability. *Remote Sens*. **2014**, *6*, 12275–12308; doi:10.3390/rs61212275.

Mishra, N., Helder, D., Angal, A., Choi, J. and Xiong, X. Absolute Calibration of Optical Satellite Sensors Using Libya 4 Pseudo Invariant Calibration Site. *Remote Sens.* **2014**, *6*, 1327-1346; doi:10.3390/rs6021327

Smith, D.L. and Cox, C.V. (2013) (A)ATSR solar channel on-orbit radiometric calibration. Geoscience and Remote Sensing, IEEE Transactions on, 51, 1370-1382.

*Willson, R.,C. (2014-05-16). "ACRIM3 and the Total Solar Irradiance database". Astrophysics and Space Science **352**: 341–352.* Bibcode:2014Ap&SS.352..341W. doi:10.1007/s10509-014-1961-4