VIIRS DNB SDR Algorithm Improvements

Steve Mills
NOAA STAR/ERT
9 August 2016

Subtopics

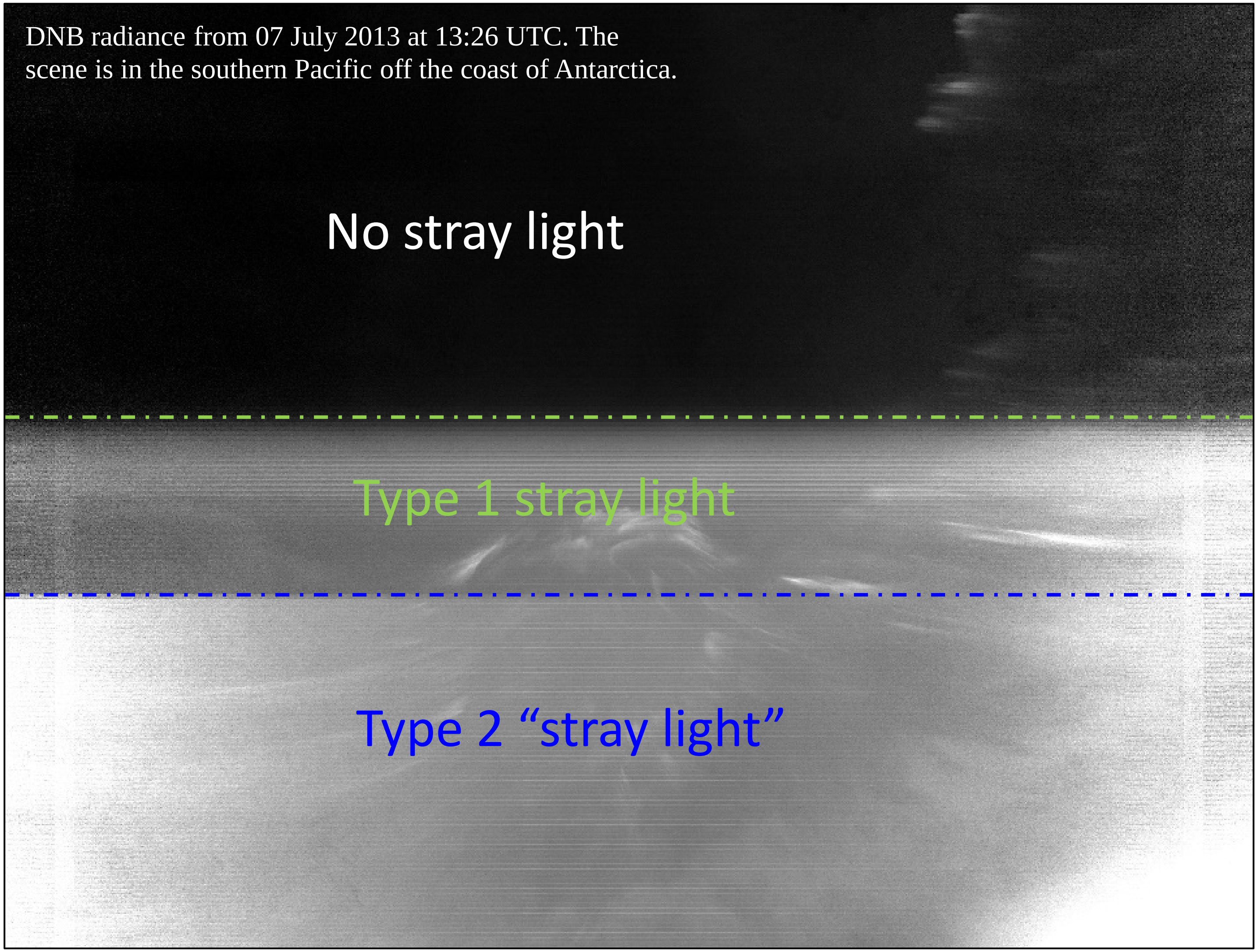
1. Cal based gain ratios and stray light correction
2. DNB Offset & Noise Analysis from Cal Data
3. Determining offsets using Earth view—a statistical method using a parametric model with method of moments estimator

Part 1 –Cal based gain ratios and stray light correction

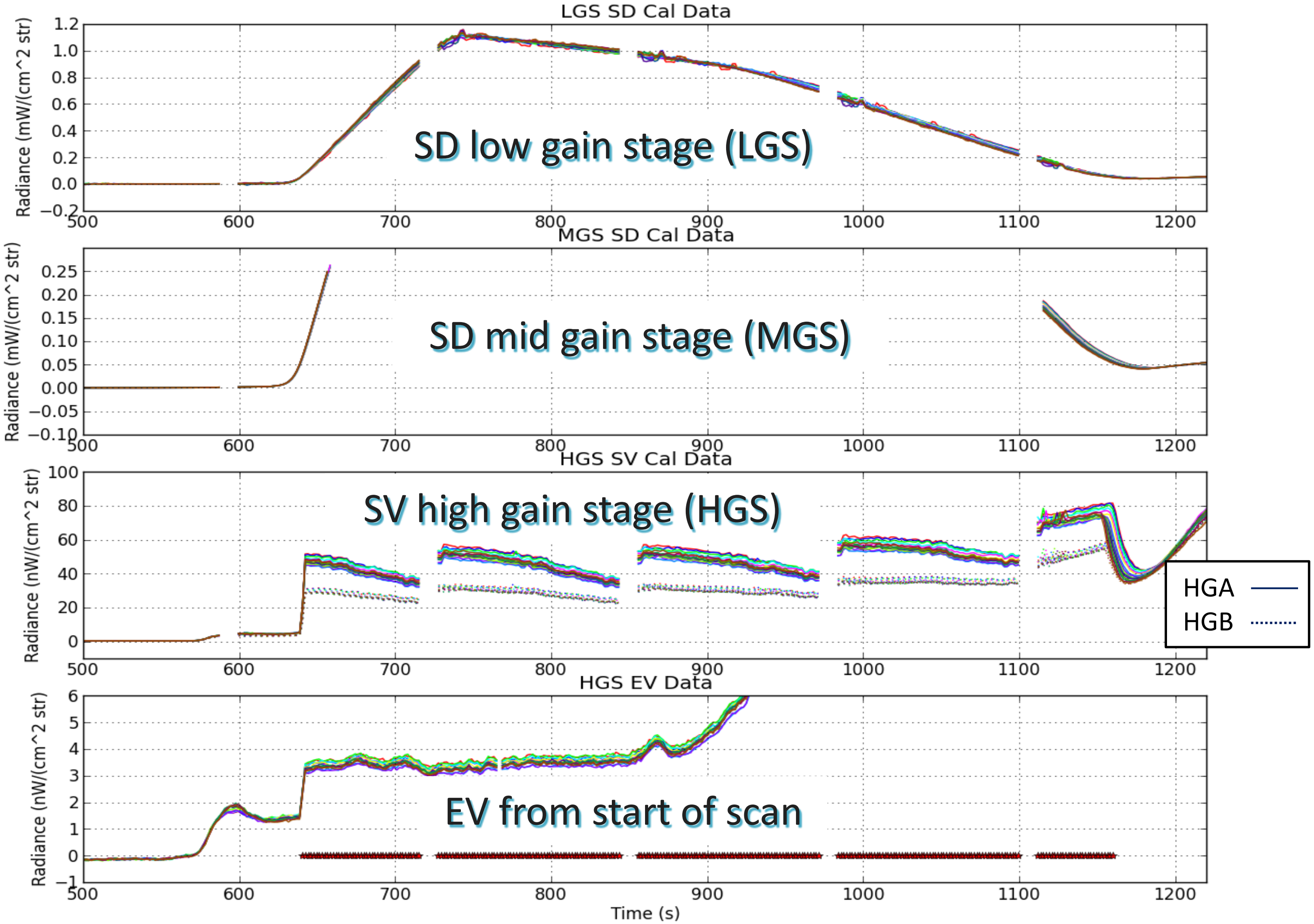
Cal Sector Data around Night to Day Transition

- **In theory the cal space-view (SV), blackbody (BB) calibration (cal) should always be dark throughout the orbit.**
  - Space has very little light except for stars and airglow in the ionosphere
  - The blackbody is black, meaning that it should not reflect any light
- **In fact, both the SV and the BB are affected by stray light**
  - They have strong signals during the day and around both terminator crossing.
  - This stray light is correlated with the stray light seen in the earth-view (EV)
- **The stray light has been shown to be dependent on satellite solar zenith angle (SSZA) and satellite solar azimuth angle (SSAA)**
  - during the night to day transition in the southern hemisphere the stray light is quite different from the northern hemisphere because of the SSAA is in the opposite direction.
  - There is a seasonal change in the SSAA over the orbit and so the stray light changes from month to month.
- **The solar diffuser (SD) cal sector data is almost always the strongest**
  - This is not surprising since the SD has almost 100% reflectance
  - Even when the sun is not directly illuminating the SD it apparently is being illuminated by earthshine throughout the twilight, daytime and even nighttime.

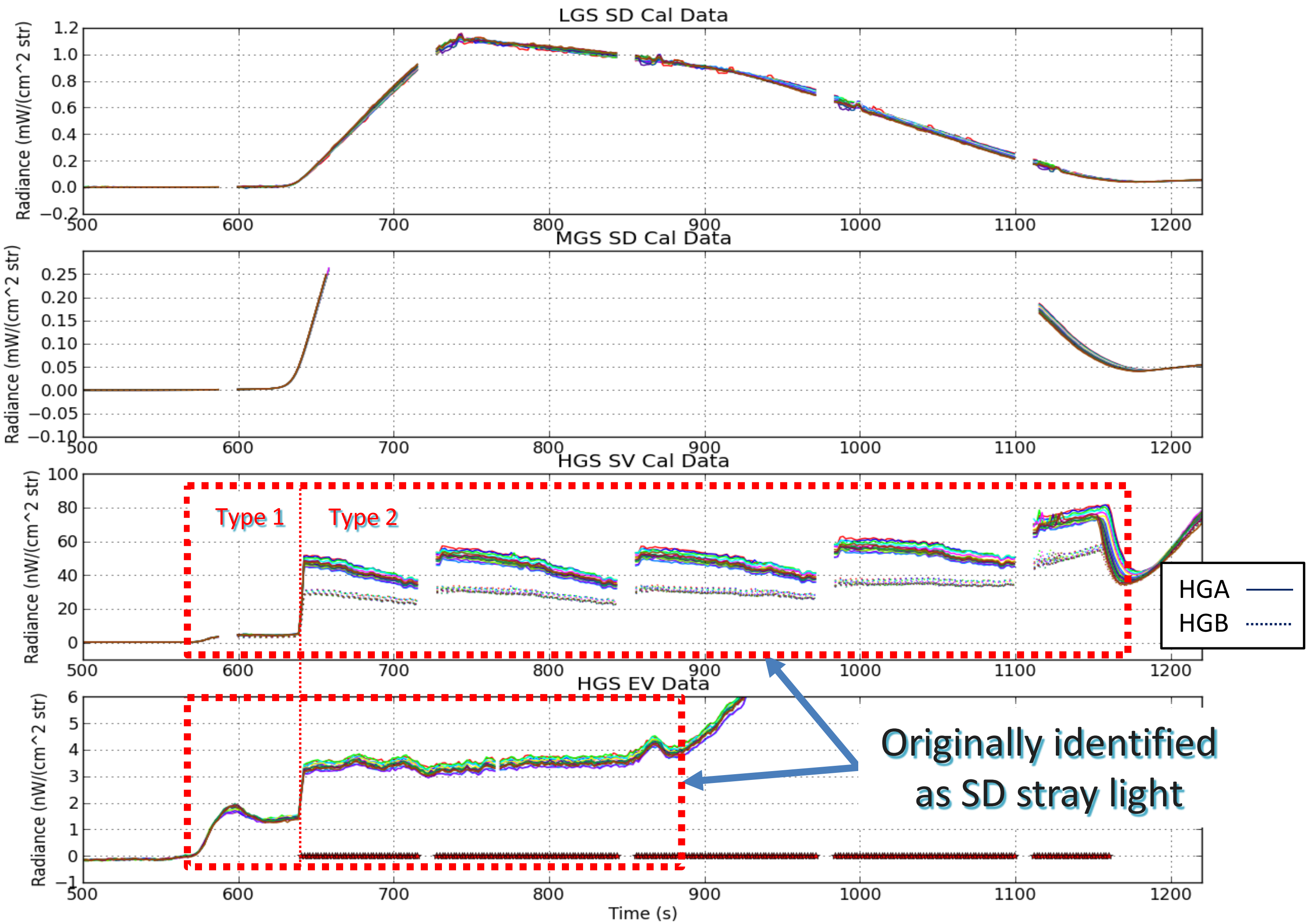
EV “Stray light” around night-to-day terminator crossing



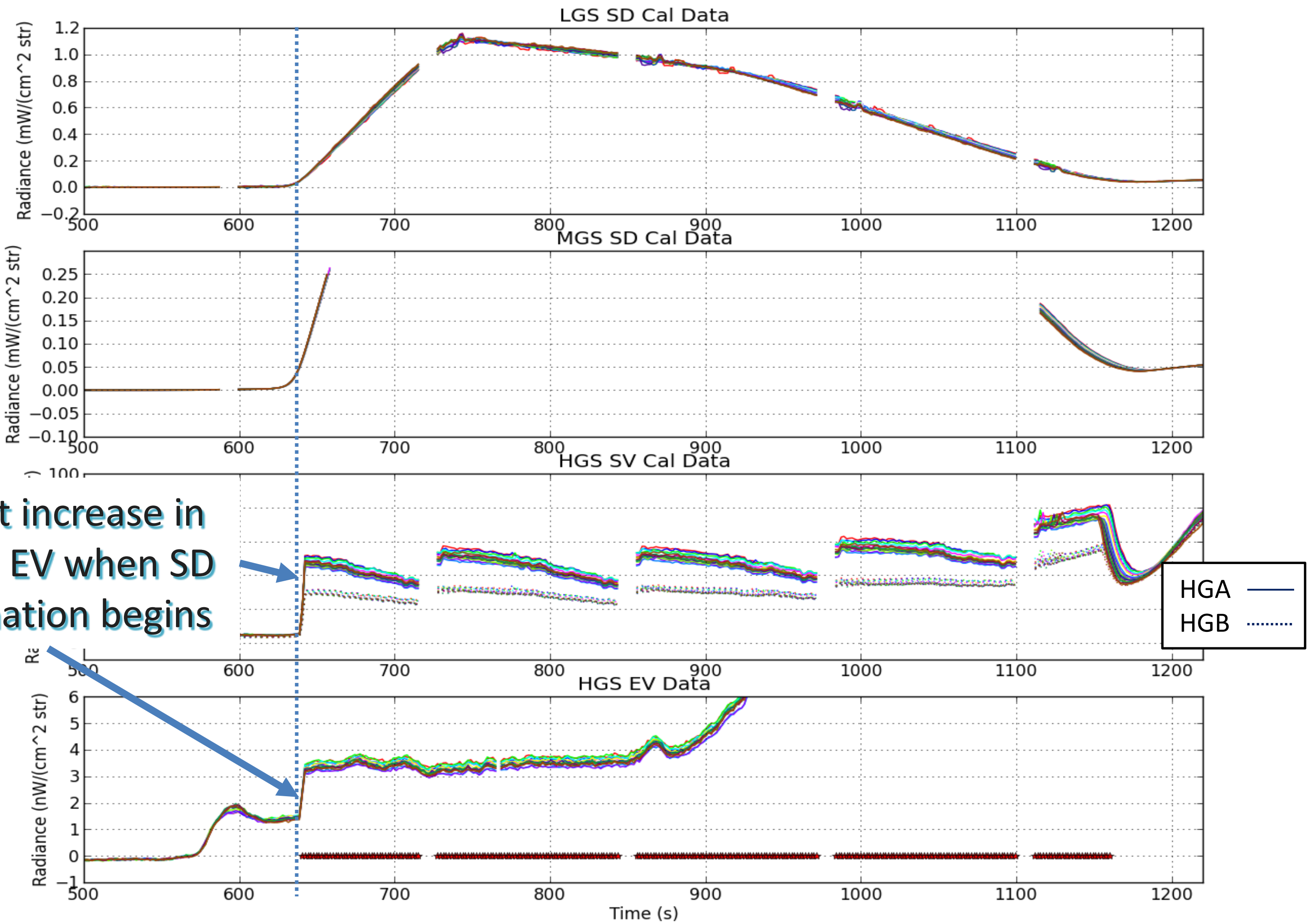
Comparing SD, SV and EV signals during night to day transition in southern hemisphere



Comparing SD, SV and EV signals during southern hemisphere terminator crossing

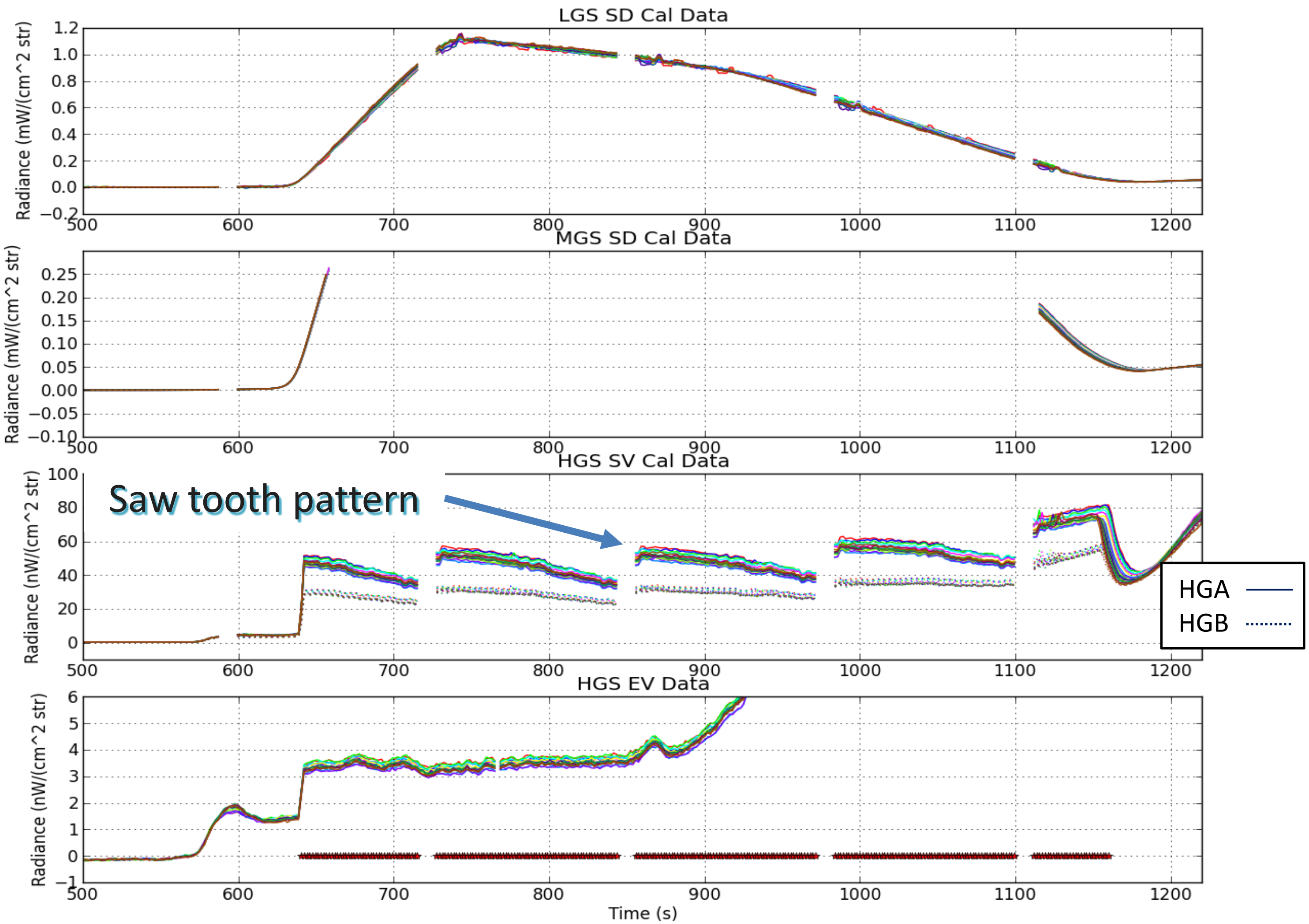


Comparing SD, SV and EV signals during southern hemisphere terminator crossing

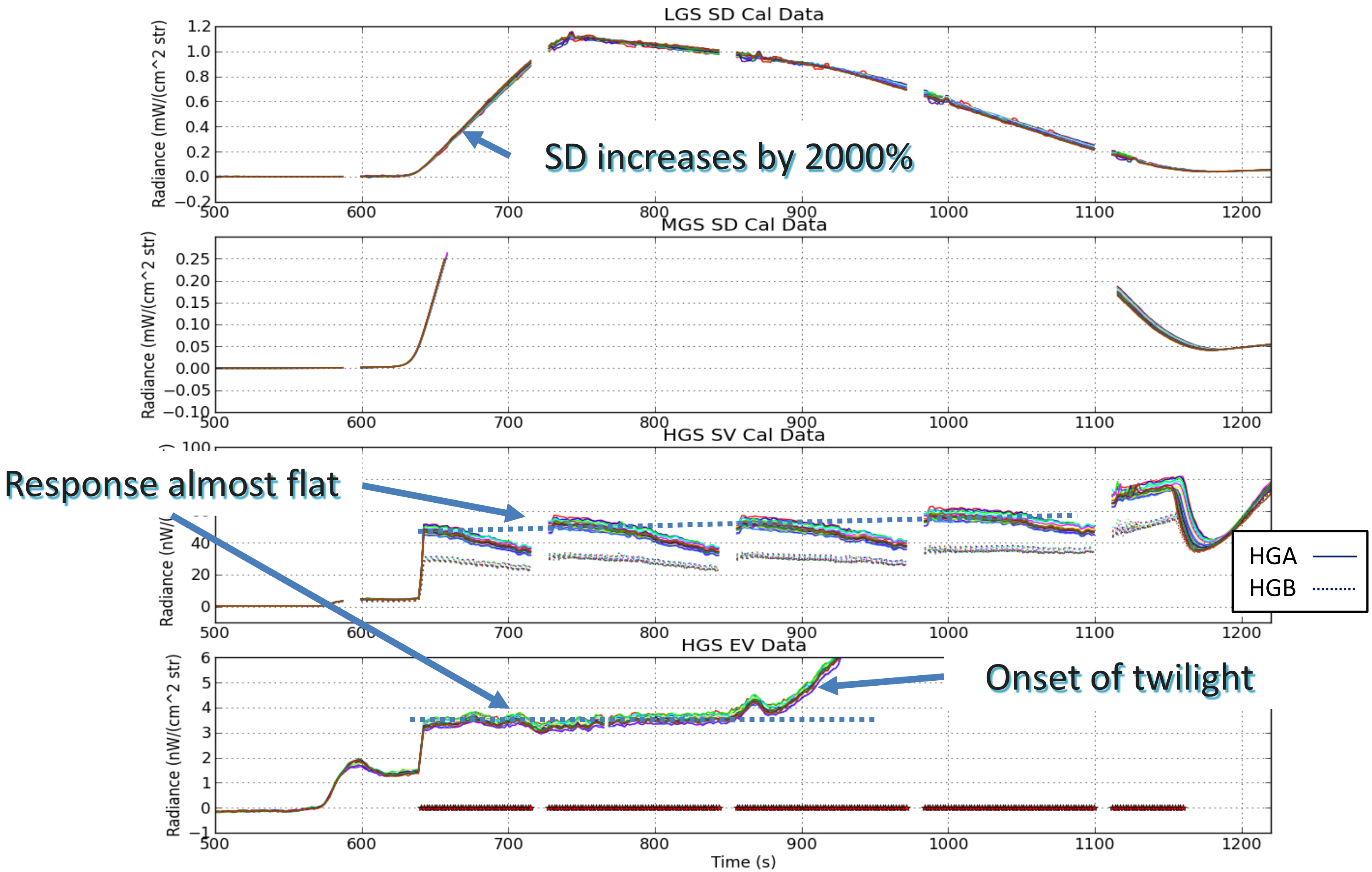


Abrupt increase in SV and EV when SD illumination begins

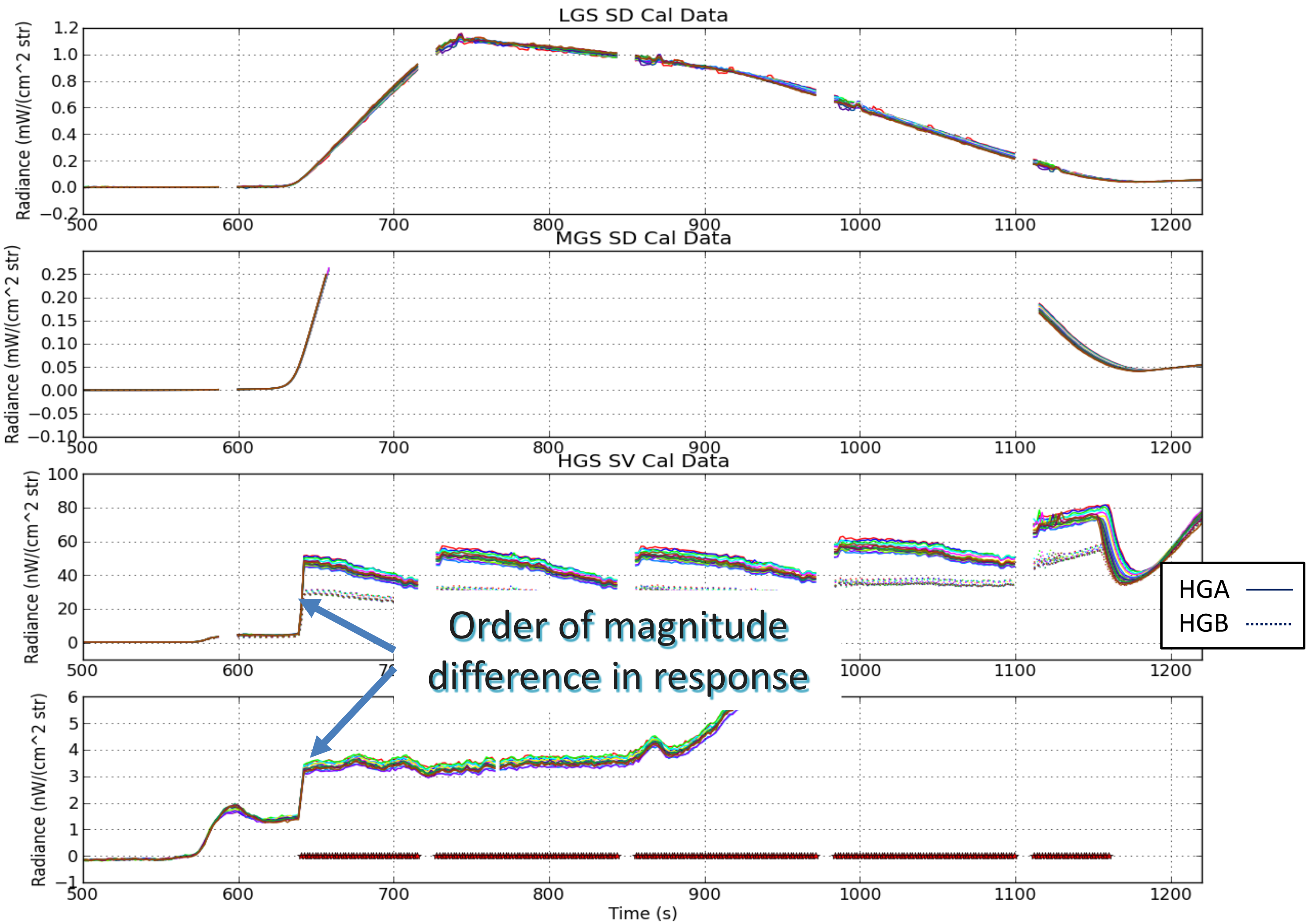
Comparing SD, SV and EV signals during southern hemisphere terminator crossing



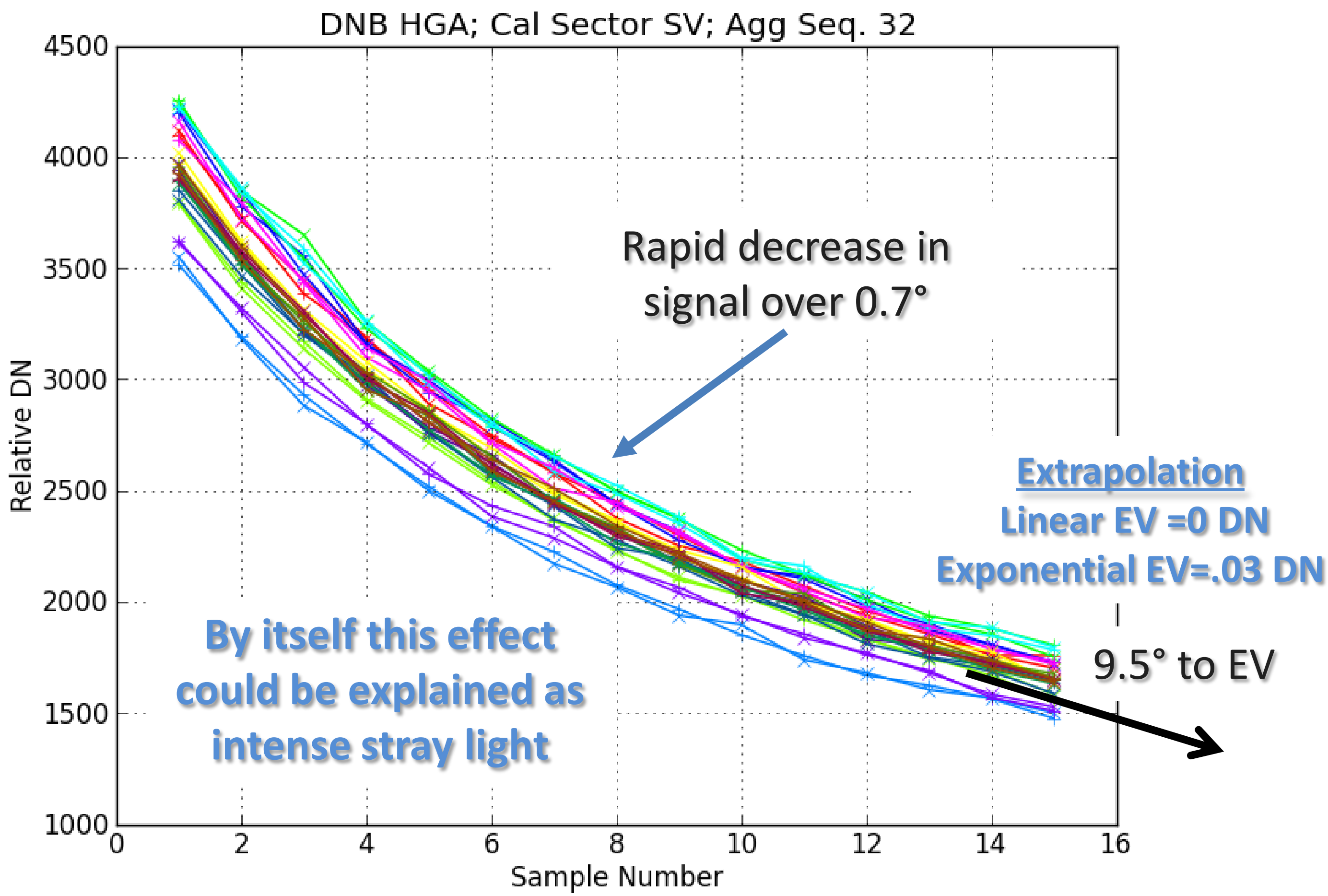
Comparing SD, SV and EV signals during southern hemisphere terminator crossing



Comparing SD, SV and EV signals during southern hemisphere terminator crossing



SV signal during this period

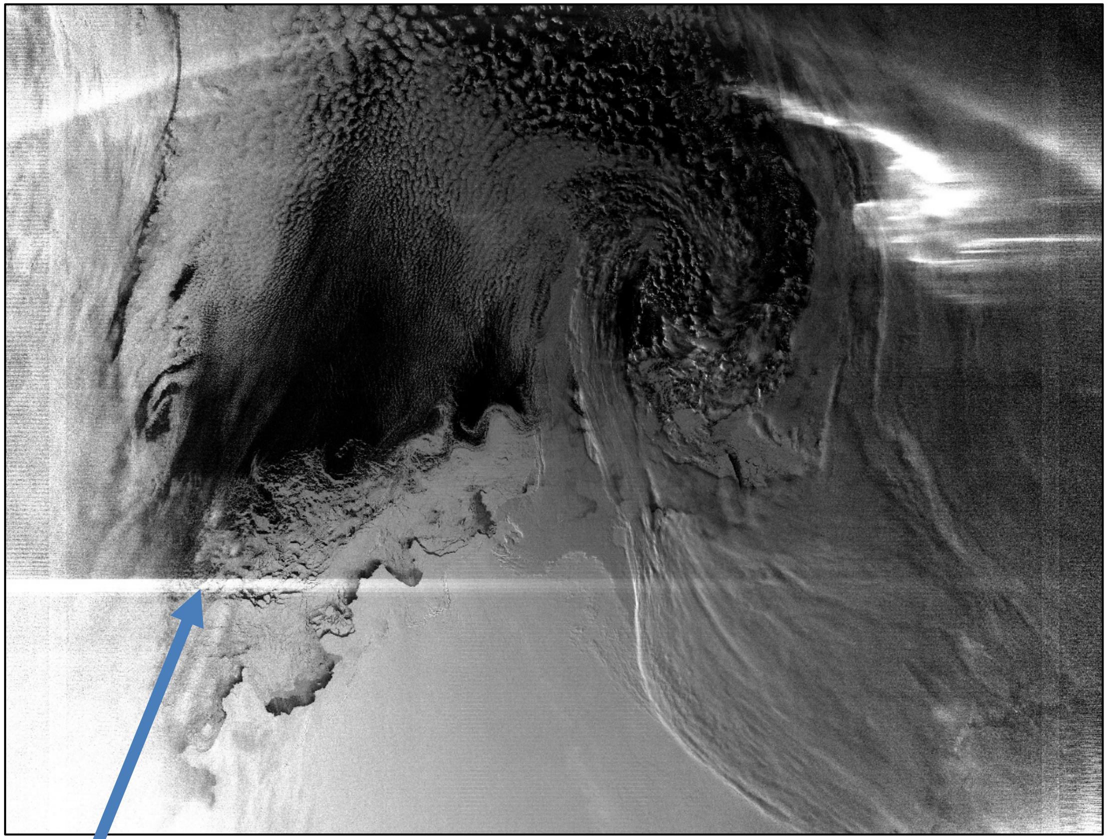


Inconsistencies with stray light hypothesis—What is really going on?

- **Alternative hypothesis**—hysteresis from intense SD signal
  - There are known to be problems with the S-NPP VIIRS DNB anti-blooming electronics
  - Other mechanisms are possible, e.g. some super-saturation effect in the HGS CCD
- **Fact 1:** HGS SV & EV signals abruptly start and end over the direct solar illumination of the SD
  - **Possible cause:** anti-blooming circuit abruptly triggered with rapid increase in SD radiance
- **Fact 2:** Saw tooth pattern on HGS SV related to aggregation mode
  - **Possible cause:** because the SV signal rapidly decreases & the with lower aggregation there is less time per sample, the overall decrease is less with less aggregation
- **Fact 3:** SV & EV HGS signals otherwise uncorrelated with SD signal but instead have a flat response
  - **Possible cause:** anti-blooming trigger causes an excess charge that is fixed and thus causing offsets
- **Fact 4:** EV signal is order of magnitude les than SV mean
  - **Possible cause:** Excess charge is rapidly discharged for every sample of SV and continues after the 16 samples that are transmitted

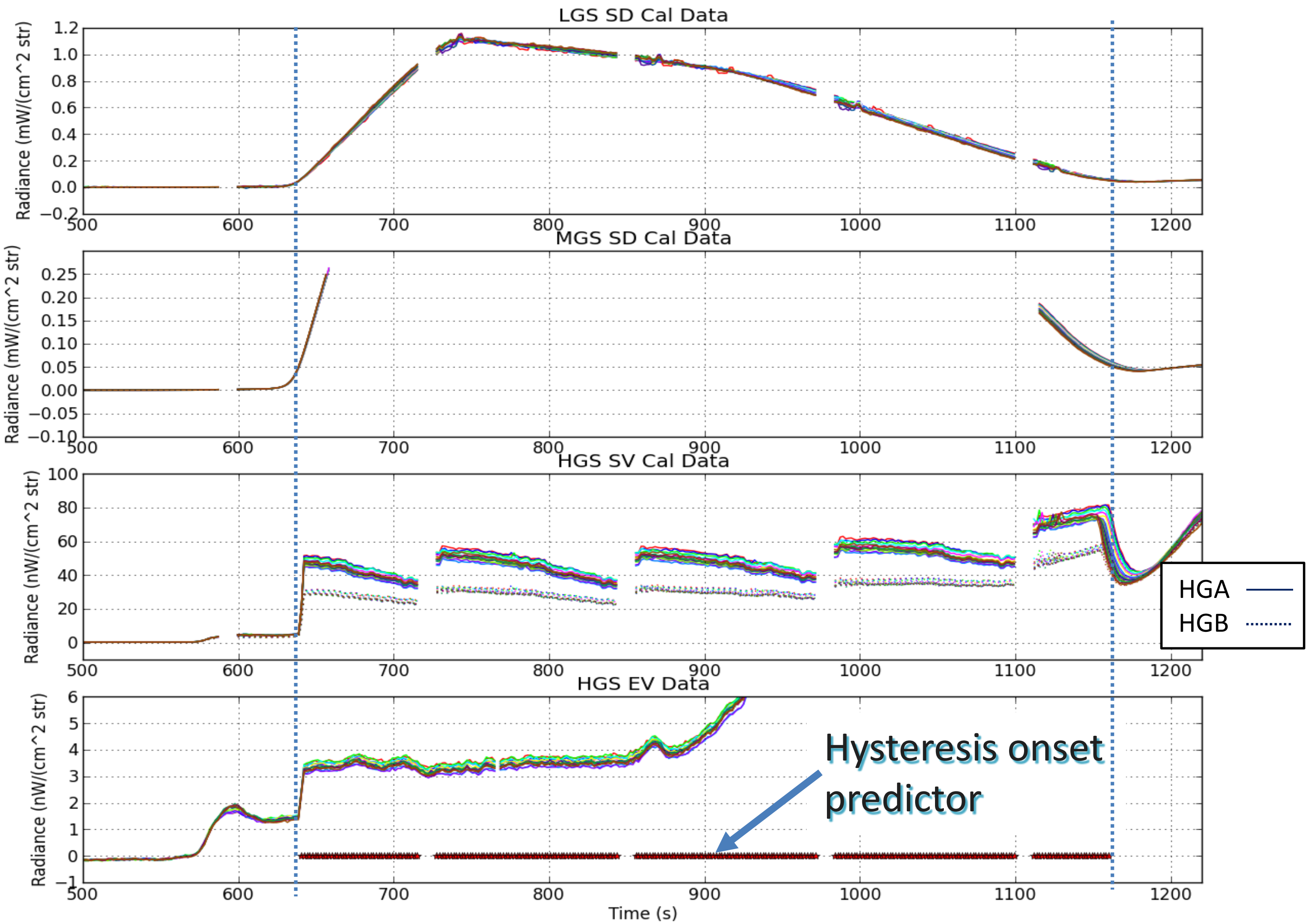
How does this effect “stray light” correction?

- The type 2 “stray light” offset correction works well over most of the area.
- In the current algorithm the onset of type 2 stray light uses the solar angle with respect to the satellite.



- The onset prediction often is off by up to 3 scans, leaving either a dark or light streak from under-correction or over-correction.
- A better predictor would be to use the SD signal with a threshold. The threshold of 0.06 mW cm^{-2} $steradian^{-1}$ was used to test this.

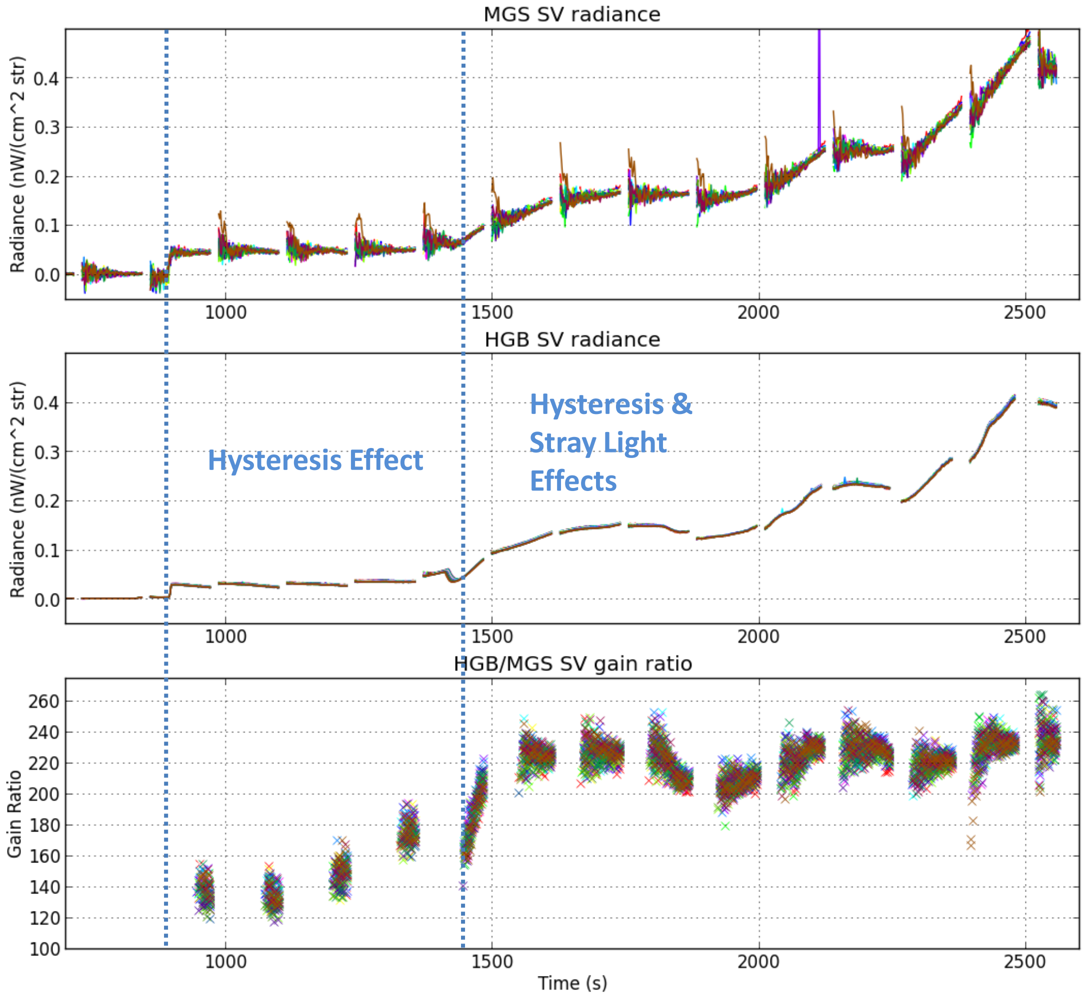
Using a SD threshold to predict the SV and EV onset



Hysteresis affect on gain ratios from SV or BB

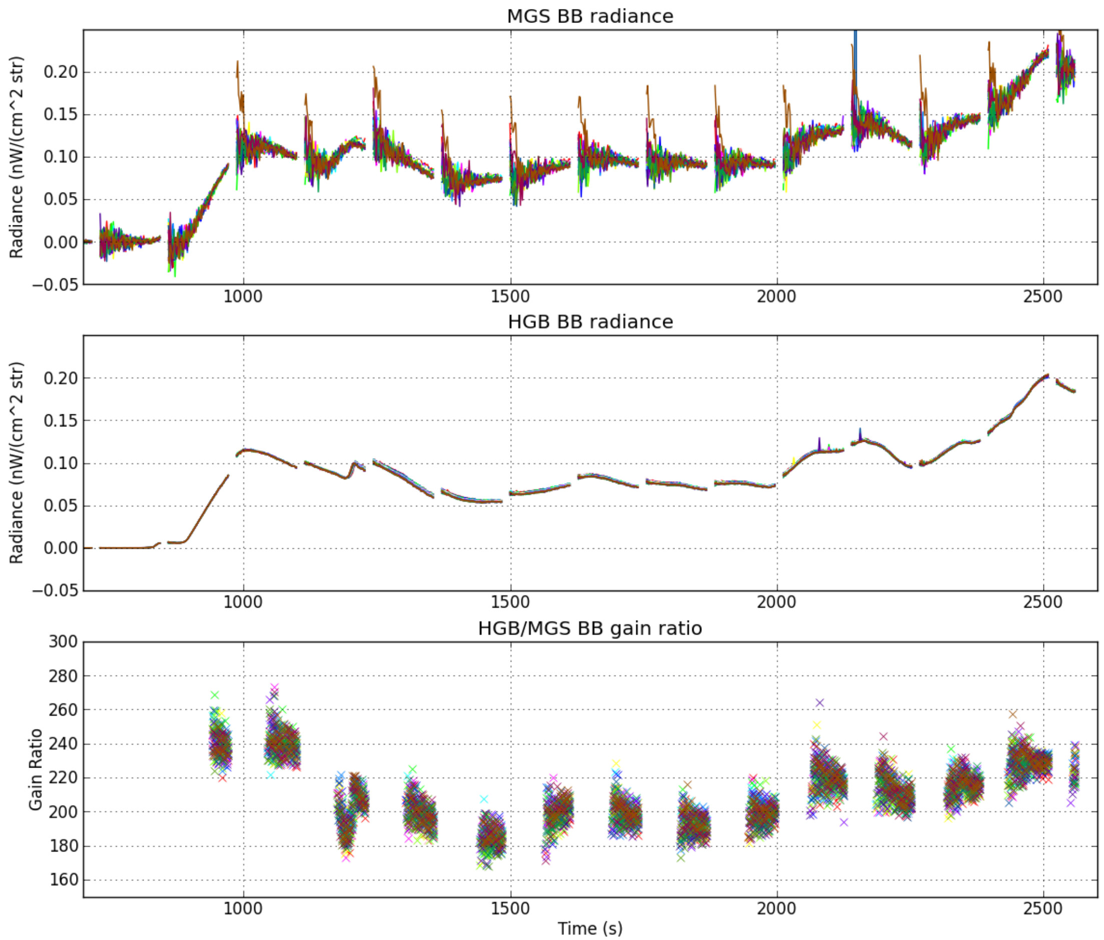
- For the SD the HGA & HGB signals are almost always saturated and it is rare that MGS/HGS gain ratios can be produced (4 to 5 scans per orbit).
- Therefore RSBAutoCal currently uses MGS/HGS gain ratios from the SV and BB signal.
- Since the SV & BB should ideally have no signal, almost all the observed signal must be stray light or hysteresis.
- The validity of these gain ratios is based on the assumption that the stray light equally illuminates both MGS and HGS.
  - There is evidence that this is not true, even on average.
- During the daytime the EV signal is as strong as or stronger than the SD signal during solar illumination.
- The BB HGS signal is therefore likely to also be affected by hysteresis during daytime.
- If some of the BB or DV signal is hysteresis then it is not even optical, so this further invalidates using SV or BB for gain ratios.
- Therefore, the SV and BB signals should never be used for cross-calibration.
  - Automatic cross-calibration for MGS/HGS is therefore not possible.
  - Automatic cross-calibration using the SD for LGS/MGS gain ratios should still be effective

SV MGS/HGB cross-calibration



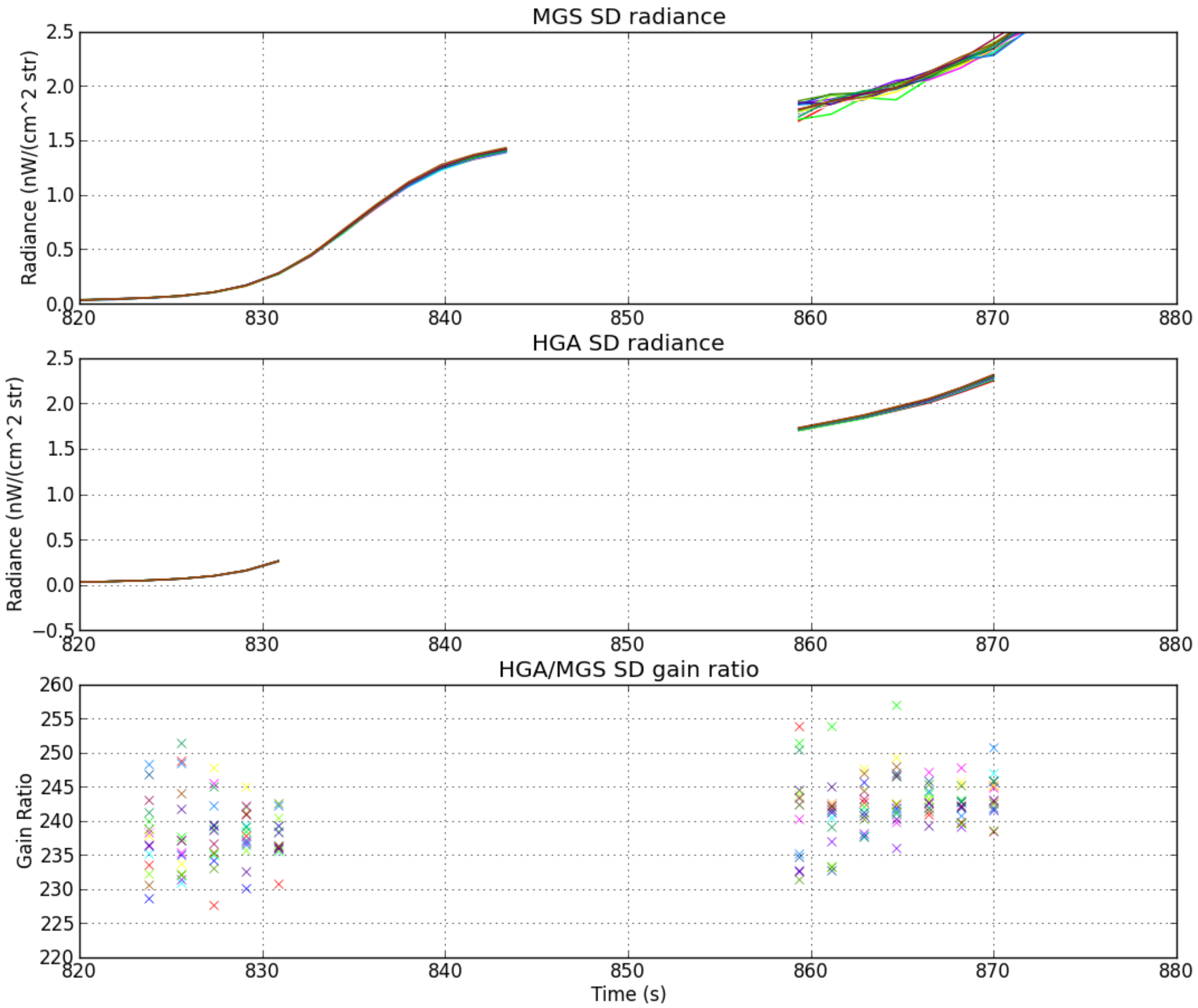
Gain ratio threshold
SNR(MGS)>15

BB MGS/HGB cross-calibration



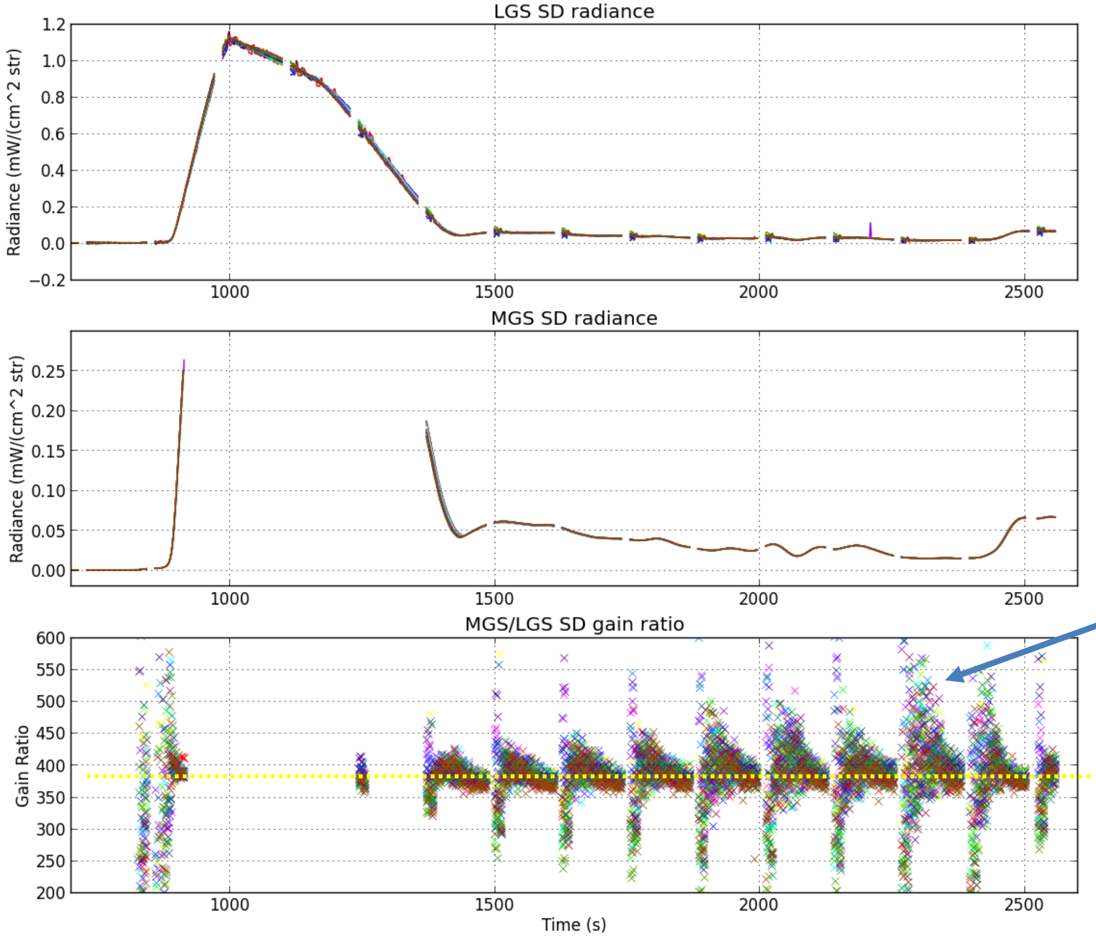
Gain ratio threshold
SNR(MGS)>15

SD MGS/HGA cross-calibration



Gain ratio threshold
SNR(MGS)>15

SD LGS/MGS day-to-night cross-calibration



Spread due to noise & nonlinearity

Stable over orbit

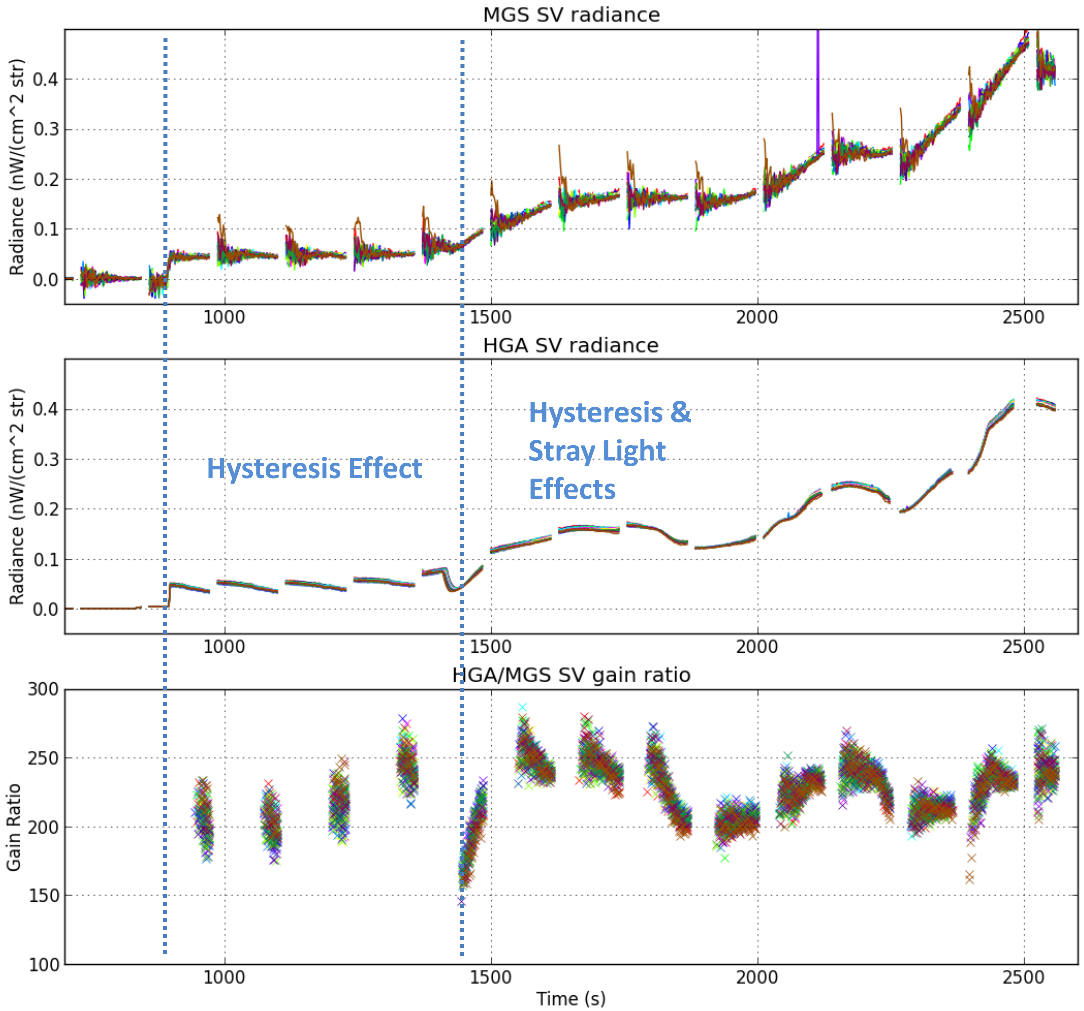
Gain ratio threshold SNR(LGS)>15

Conclusions & Recommendations

- What was previously thought to be stray light from direct sunlight on the SD is actually a hysteresis effect
  - The cause is unknown but it may be related to anti-blooming in the HGS
- The hysteresis affects the SV cal signal about an order of magnitude more than the EV
  - It rapidly decreases over the 16 cal samples
  - The rapid decrease explains a saw tooth pattern in the SV that is associated with the 72 scan DNB cal cycle
- The onset of the effect for the SV and EV is abrupt after which it immediately goes to a flat response
- Prediction of the onset has always been a problem for the stray light correction, but using a simple threshold on the SD signal onset can be predicted to within one scan.
- It is likely that the BB cal signal is also impacted by hysteresis from the EV during the daytime.
- Hysteresis adversely affects gain ratios derived from the SV or BB.
  - Large uncertainties (>50%) are probably due to hysteresis and uneven stray light illumination.
- RSBAutoCal should not use the SV or BB to produce gain ratios
- Probably there is no reliable way to automatically produce MGS/HGS gain ratios

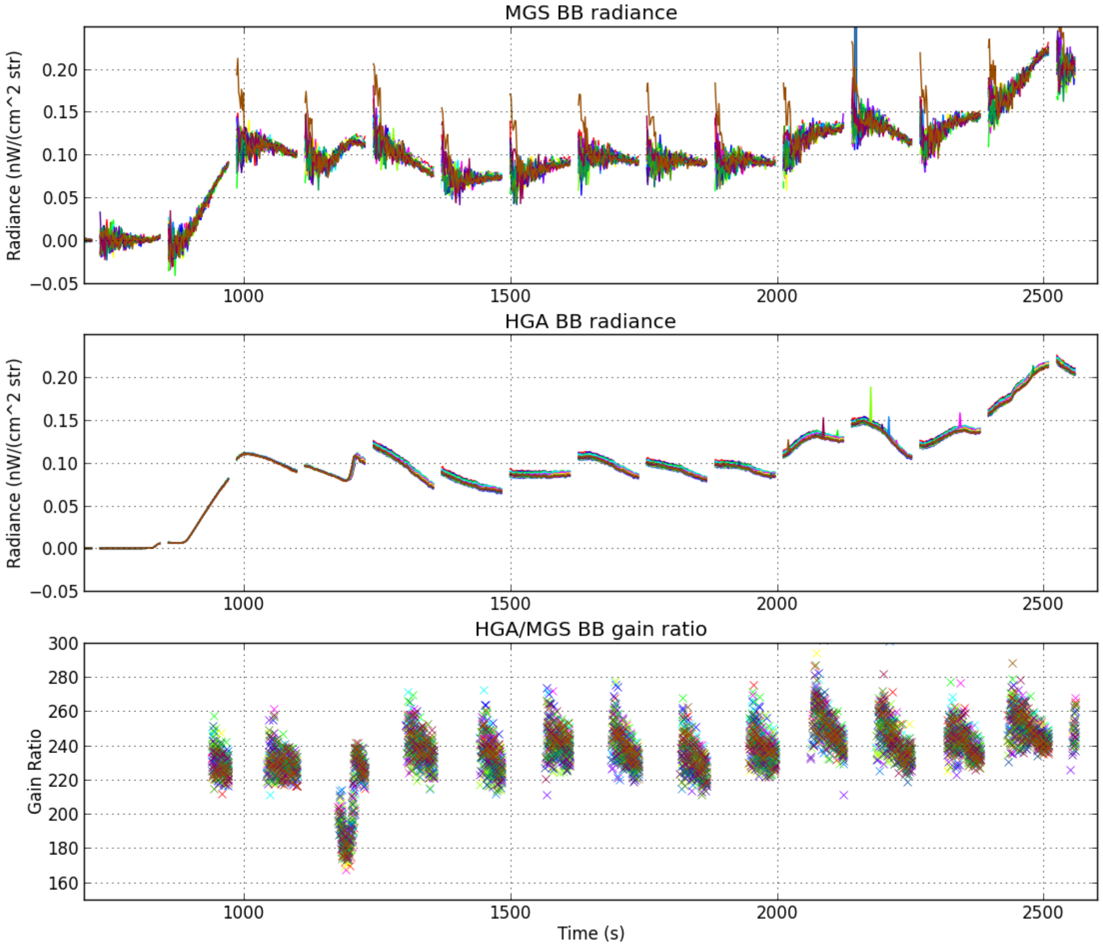
Part 1—Backup

SV MGS/HGA cross-calibration



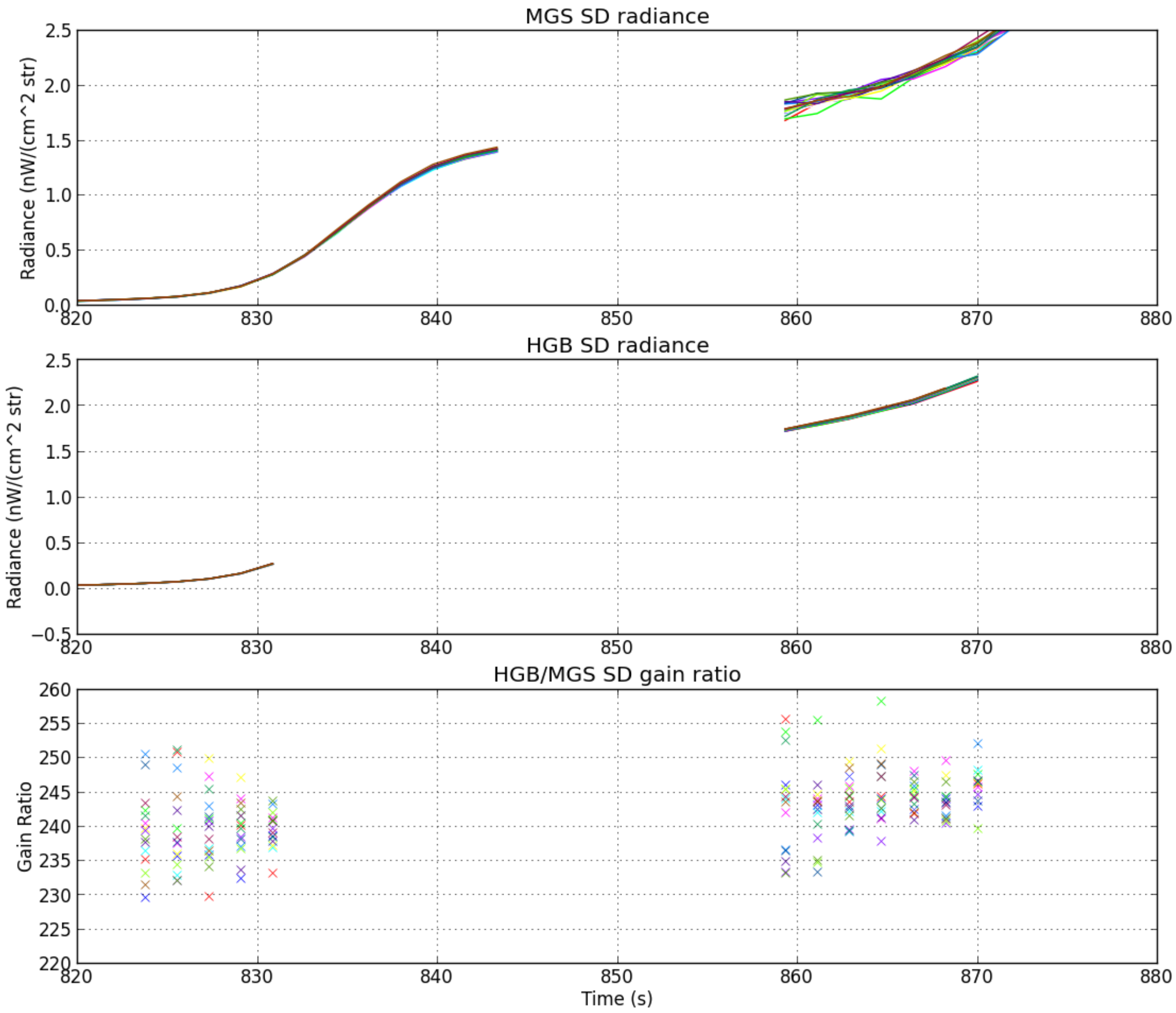
Gain ratio threshold
SNR(MGS)>15

BB MGS/HGA cross-calibration



Gain ratio threshold
SNR(MGS)>15

SD MGS/HGB cross-calibration



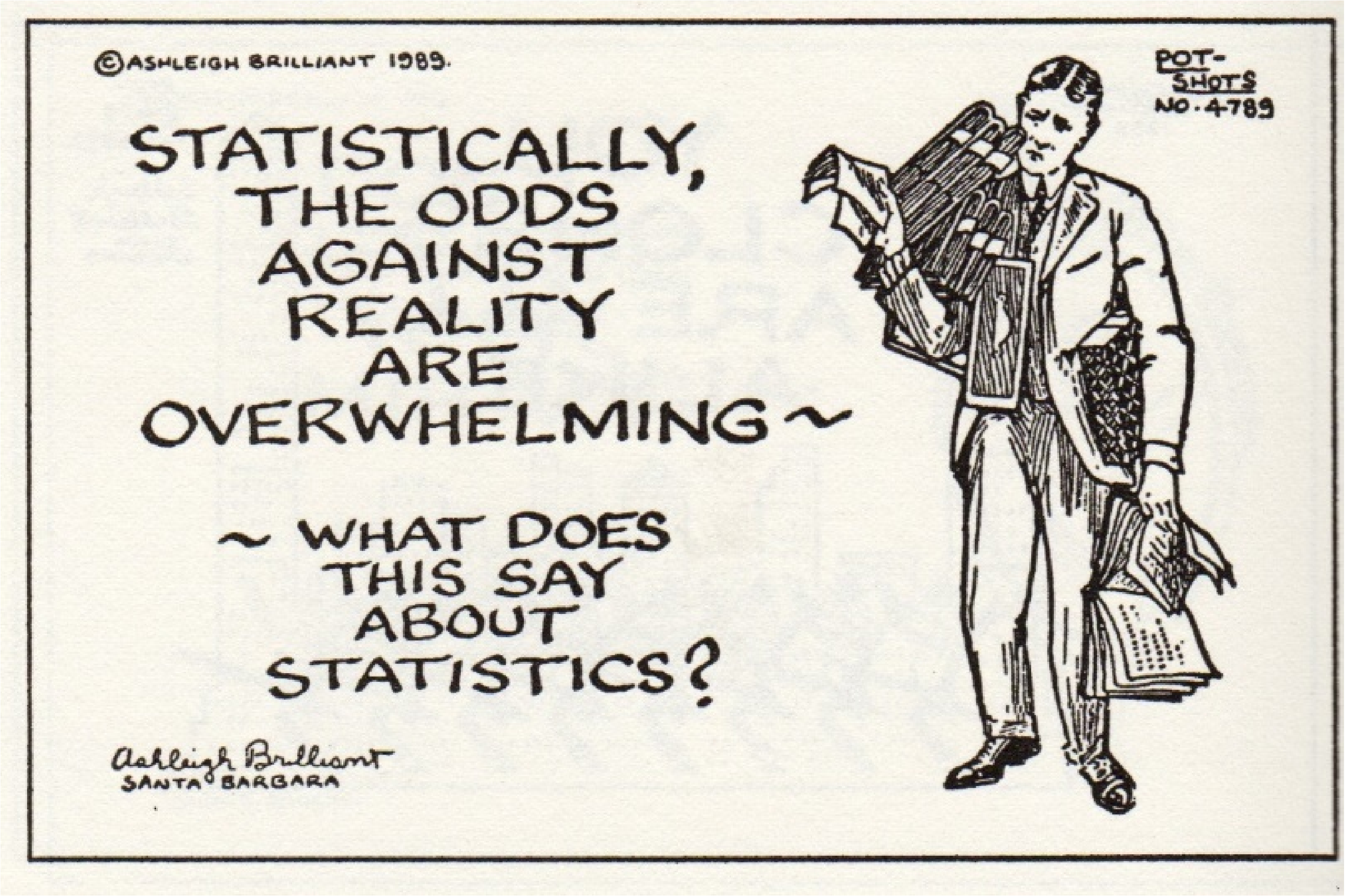
Gain ratio threshold
SNR(MGS)>15

Part 2—DNB Offset & Noise Analysis from Cal Data

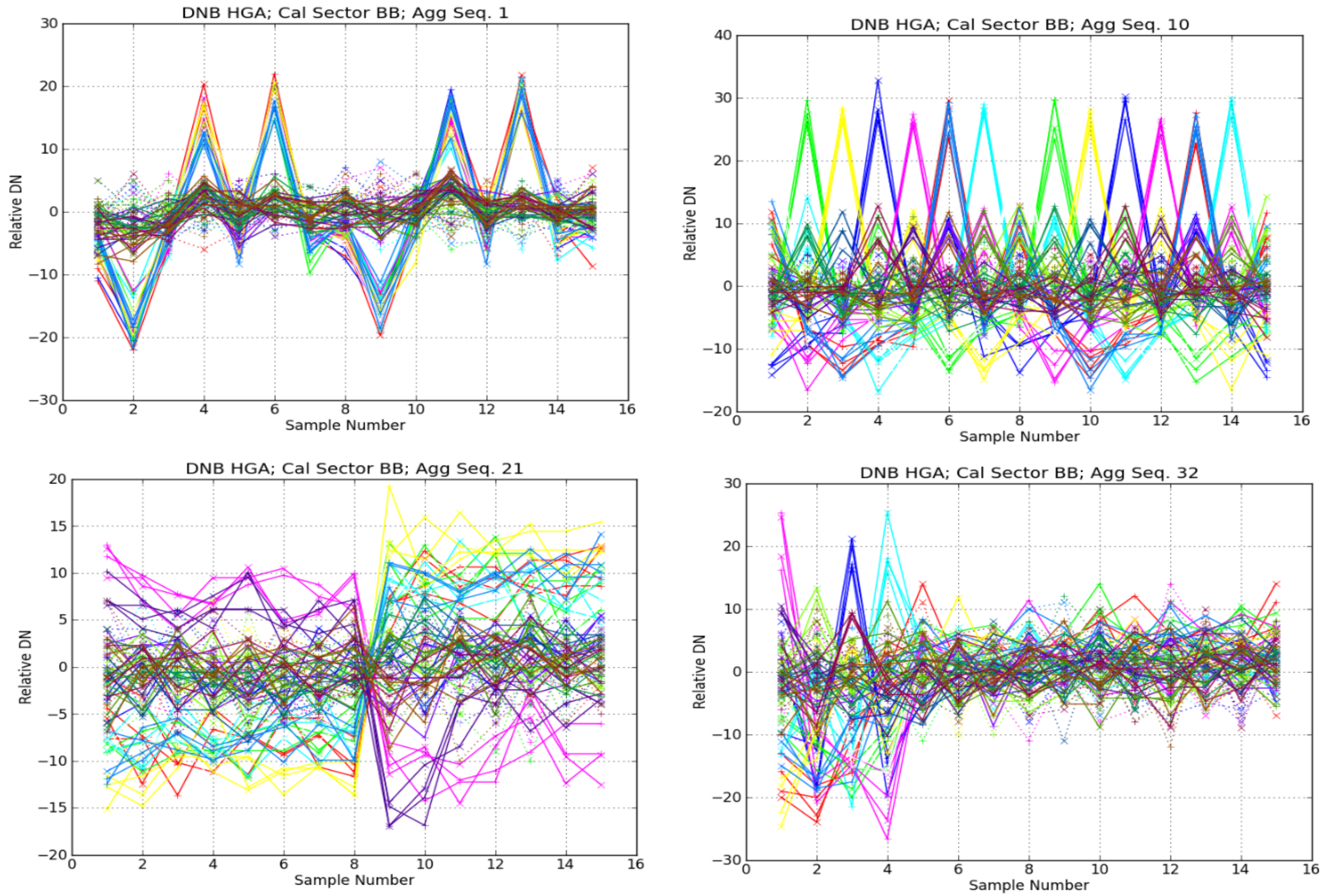
DNB Offset using cal data

- Because of fixed pattern offsets in the EV, the cal sector data cannot be used to determine absolute offset
- It has been proposed that dark cal data can be used for tracking the relative change in dark offset
  - This has not been demonstrated to be true for all agg modes
- Rather than replicate what has already been done I try to develop a <u>statistically rigorous</u> method for determining these offset along with the dark noise
- In particular, evaluation of fixed pattern offset in the cal samples and method for outlier removal are considered

Will use a statistically rigorous methodology

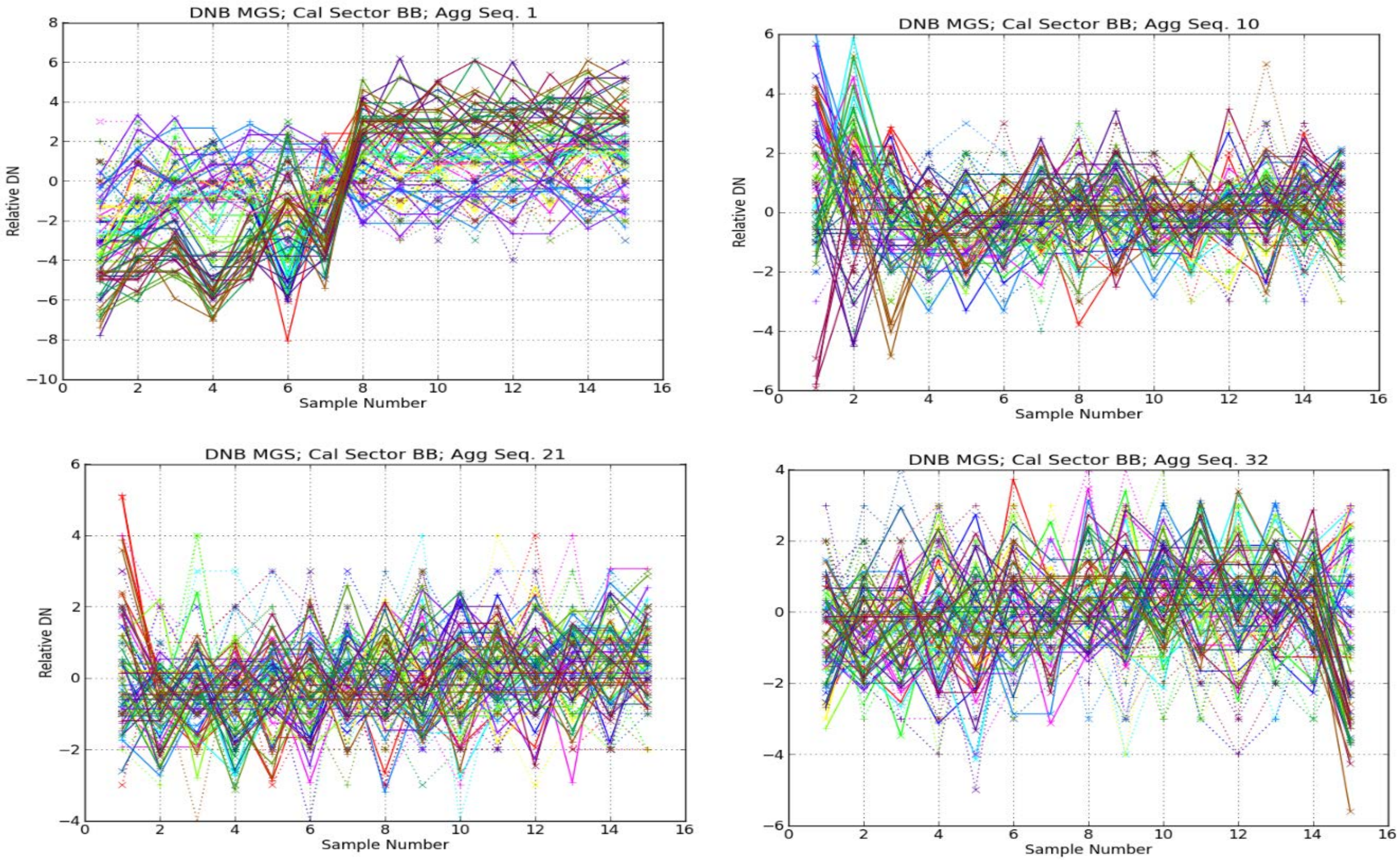


Fixed Pattern Offset (FPO) Dominates HGA Cal



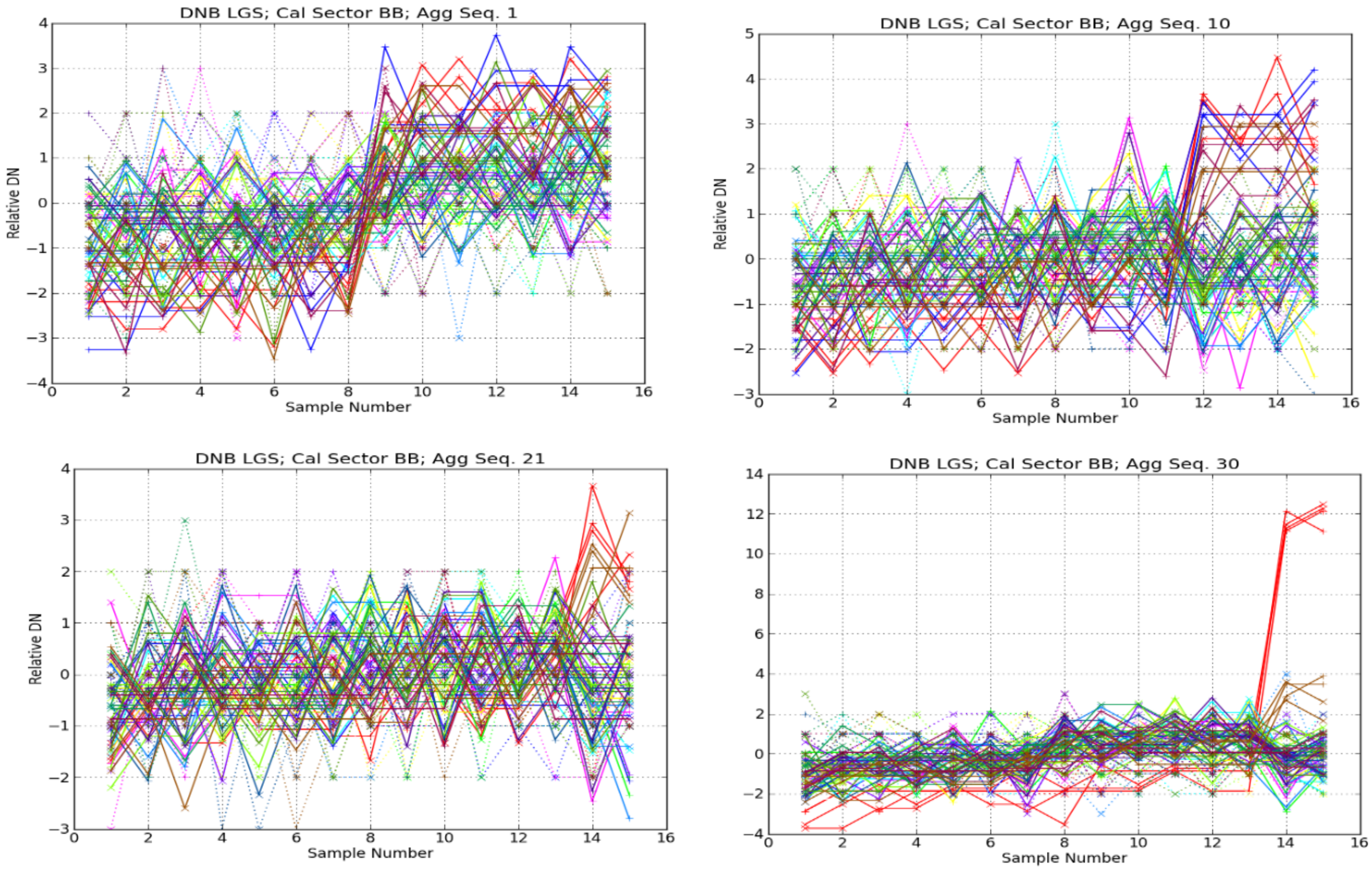
HGA Fixed Pattern Offsets for 4 aggregations. Solid lines are offsets, dotted lines are differences plotted to show noise level. Both HAM sides are plotted with "+" for side 0 and "x" for side 1. Color indicates detector number.

Fixed Pattern Offset Dominates MGS Cal



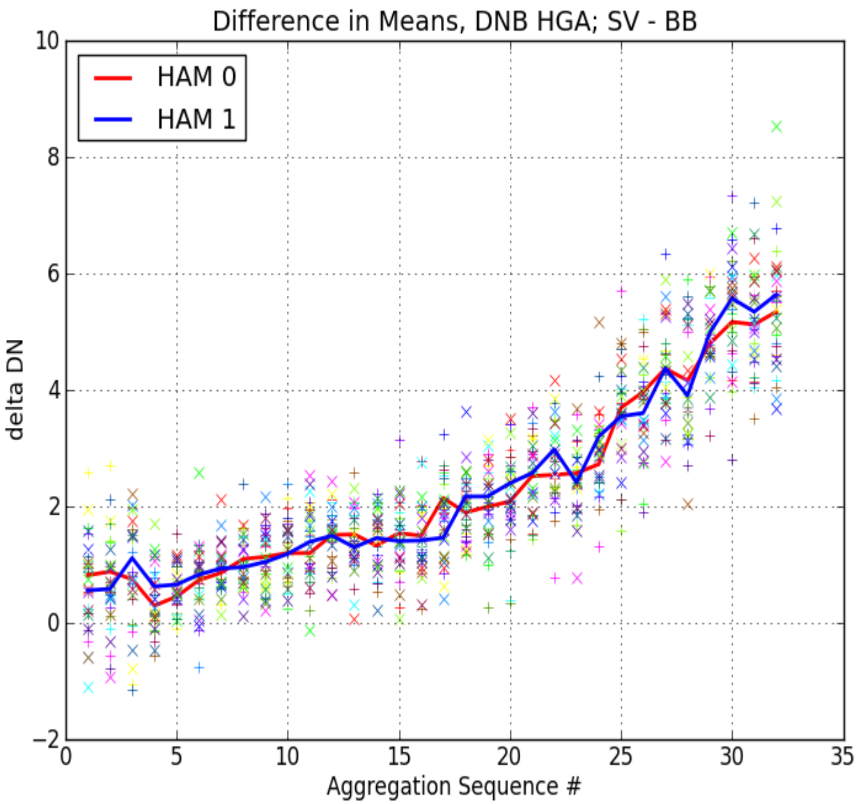
MGS Fixed Pattern Offsets for 4 aggregations. Solid lines are offsets, dotted lines are differences plotted to show noise level. Both HAM sides are plotted with "+" for side 0 and "x" for side 1. Color indicates detector number.

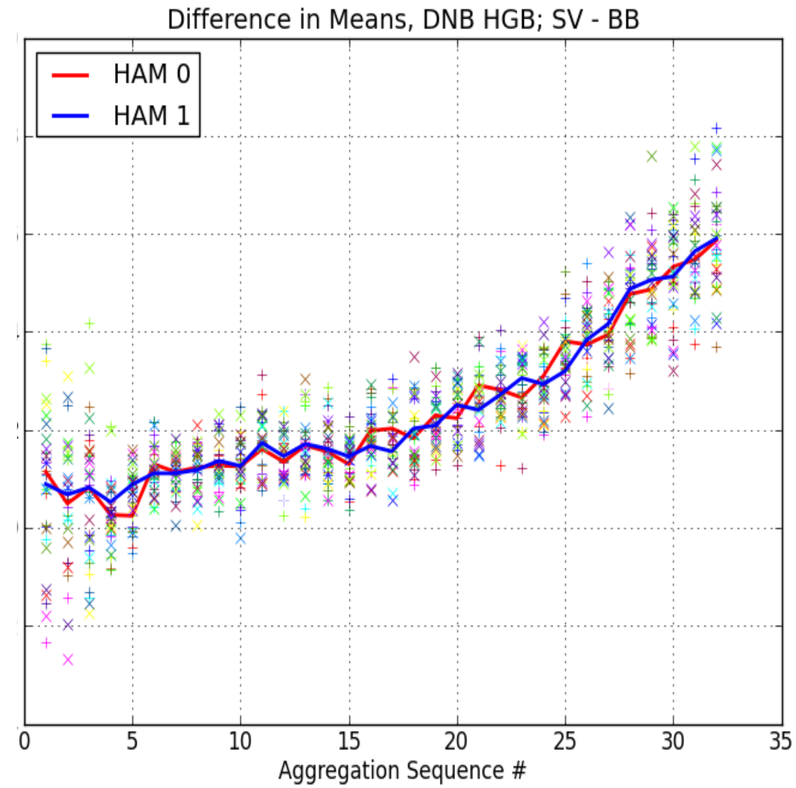
Fixed Pattern Offset Dominates LGS Cal



LGS Fixed Pattern Offsets for 4 aggregations. Solid lines are offsets, dotted lines are differences plotted to show noise level. Both HAM sides are plotted with "+" for side 0 and "x" for side 1. Color indicates detector number.

Difference in Means, HG SV-BB

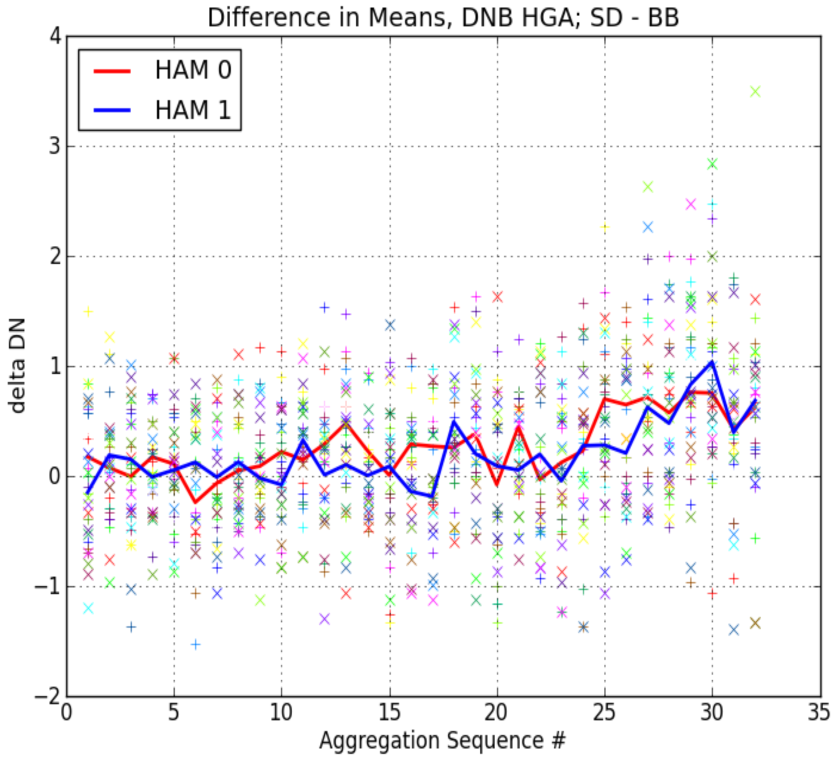


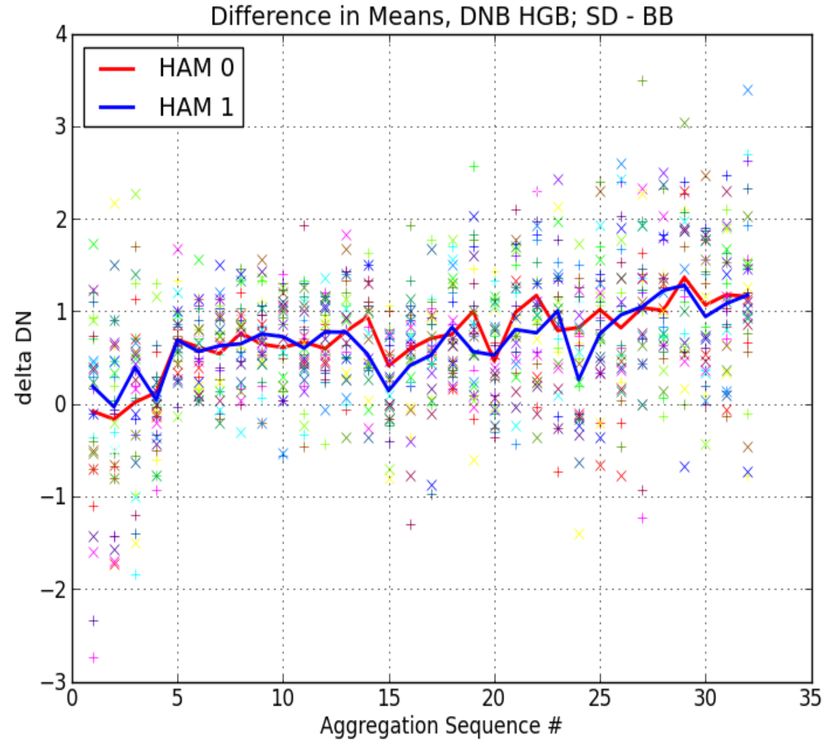


Increase due to airglow in ionosphere limb

Difference in Means, HG SD-BB

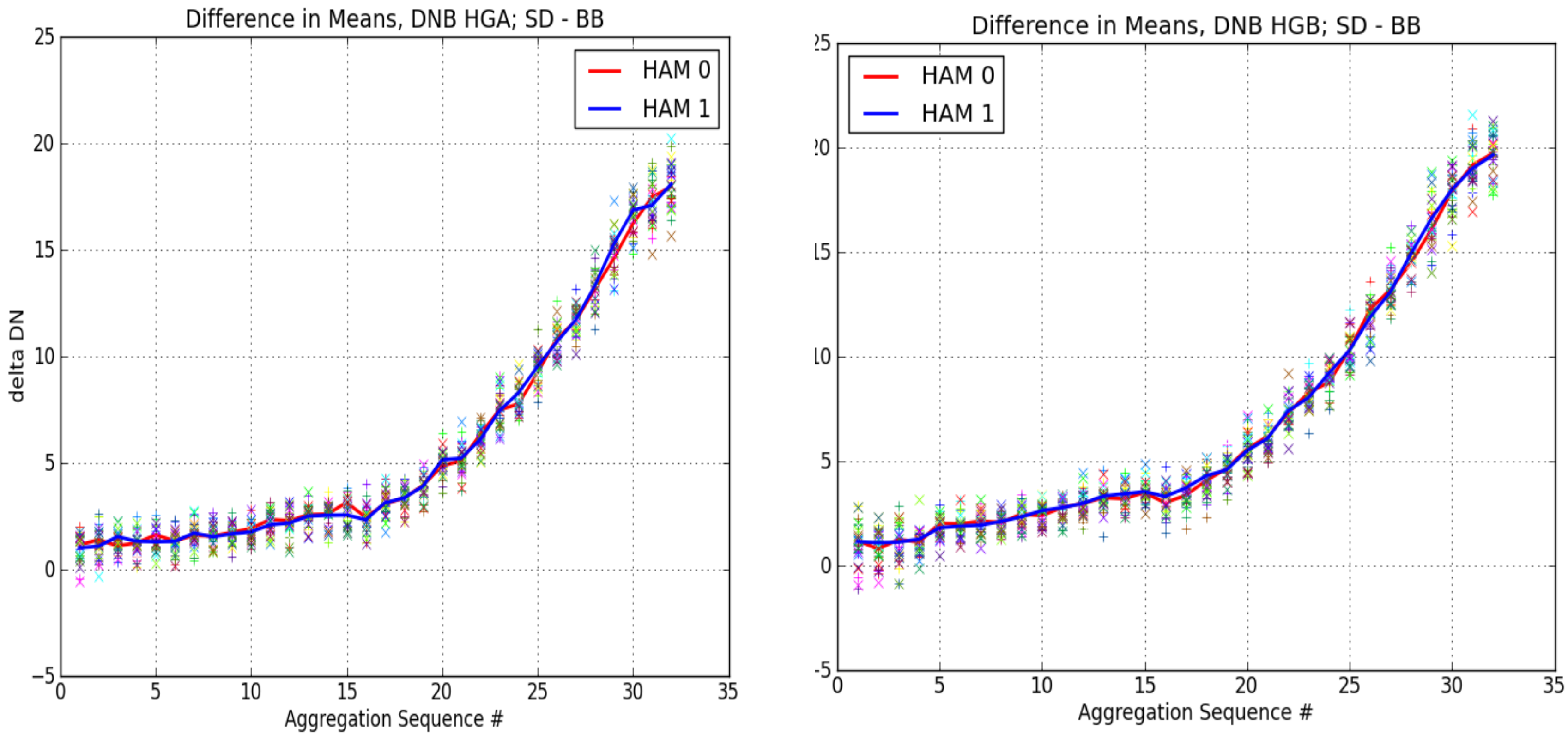
New Moon





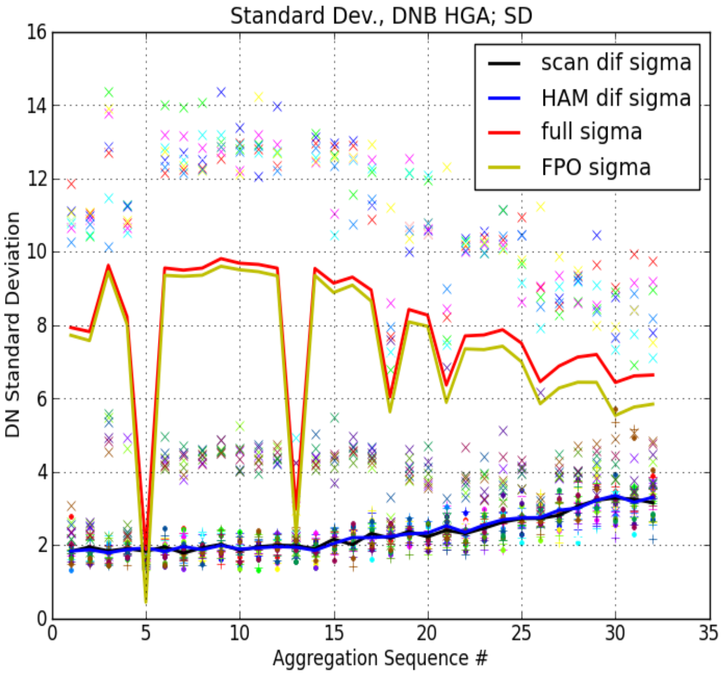
Small bias on SD due to faint light (e.g. airglow) from earth

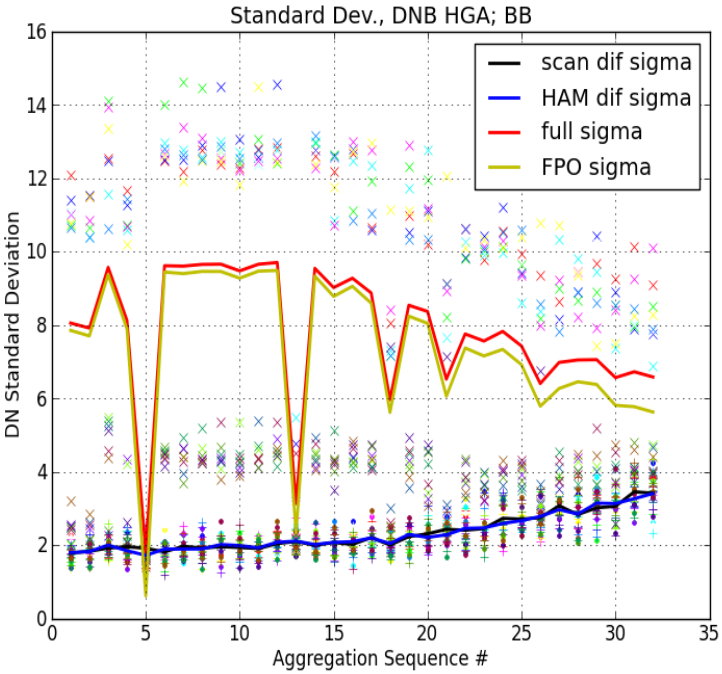
Full Moon



Large bias on SD due to moonlight from earth

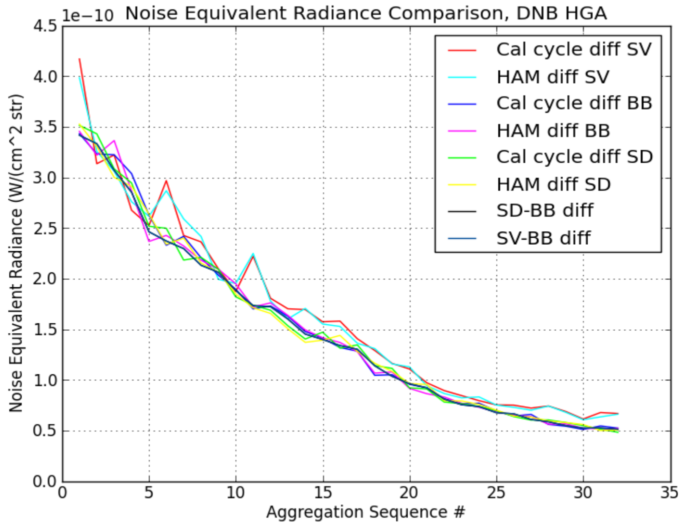
Detector noise and FPO variation compared for HGA

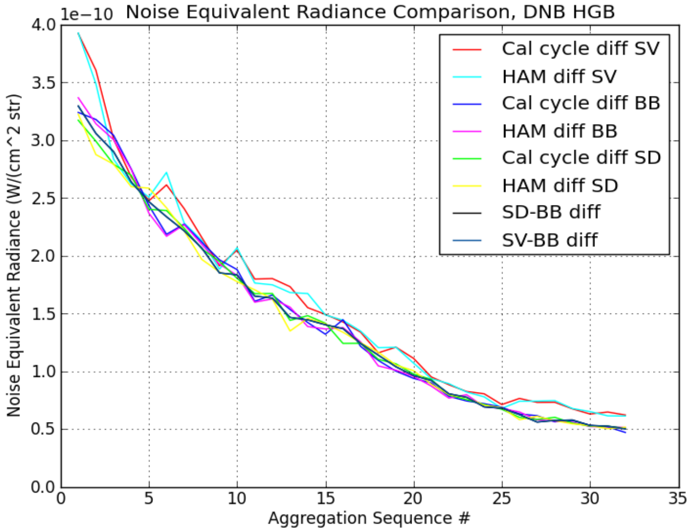


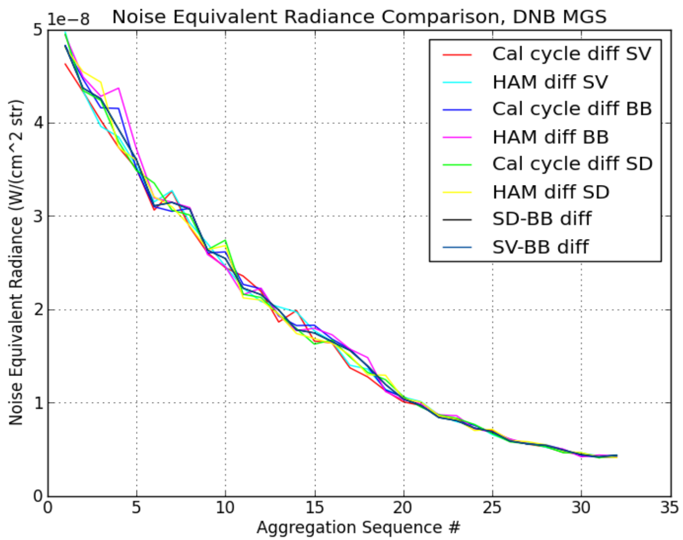


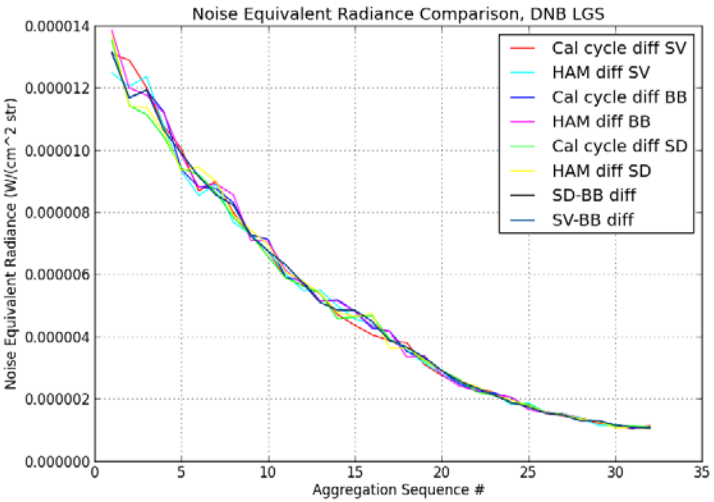
FPO variation dominates in the samples

Noise Equivalent Radiance (NER) by Aggregation









Methods of Outlier Mitigation (1 of 2)

- Winsorization –This method does not require taking the mean or the median and uses the entire ensemble to identify outliers.
  - It takes as parameters the maximum and minimum percentile of data outside of which are to be treated as outliers, for example, the lowest 2% and highest 2%.
  - Any data that is identified within the lowest range is replaced with the value exactly at the lowest cutoff point.
  - Likewise, data that is identified within the highest range is replaced with the value exactly at the highest cutoff point.
  - Because it replaces the data rather than remove it, the resulting ensemble has the same number of elements as the input ensemble.
- Trimming—This is similar to winsorization except that it removes, rather than replaces, the data below or above the percentile limits.
  - Like winsorization it takes as parameters the maximum and minimum percentile of data outside of which are to be treated as outliers,
  - but instead of replacing these values it trims these data elements from the ensemble.
  - The resulting ensemble has fewer elements than the input ensemble.

Methods of Outlier Mitigation (2 of 2)

- <u>Multilayer Median Trimming (MMT)</u>—
  - This method may be based on Peirce's Criterion which uses the median rather than the mean because outliers will always increase the standard deviation, and will shift the mean but will have very little effect on the median.
  - It is called multilayer because it repeats the trimming process multiple times with successively tighter trimming of outliers.
  - The process is to compute the median and σ, then trim values $> n\sigma$ or $< -n\sigma$ from the median.
  - This process is repeated multiple times with successively lower values of n.
  - This process is often applied with integer values of n, but this is not a necessary requirement.

Conclusions & Recommendations (1 of 3)

- The cal data may be useful in determining offset drift in the EV.
  - Because of the high uncertainty in the offsets derived from earth view data it is not certain whether in fact the drifts always correlate.
  - It has been shown that they approximately correlate for at least some detectors and Agg. Seq., but this has not been shown to be true in general.
- The BB is the best cal data to use for offset determination and for noise estimation because it is not strongly contaminated by indirect light from the earth or by airglow, unlike the SD and SV.
  - Even for the MGS, during a full moon there is sufficient light to produce a detectable offset in SD, and likewise for airglow contamination in the SV.
  - Also, including SV and SD only adds an unnecessary level of complexity to the analysis and at worst may reduce the accuracy.
- Because of FPO data ensemble averages should always be taken over the same sample over many scans
  - Per scan averages should never be taken
  - Outlier removal should be done on per sample ensembles, and never per scan

Conclusions & Recommendations (2 of 3)

- FPO is a function of sample number, detector number, Agg. Seq. and gain stage, but is not apparently a function of HAM side or cal source.
- To determine NEC without the effect of FPO, data with the same sample number, detector number, Agg. Seq. and gain stage can be subtracted, and the standard deviation taken.
- NEC determined in this way was found for:
  - HGA and HGB to go from 1.75 DN for Agg. Seq. 1 to 3.3 DN for Agg. Seq. 32.
  - MGS NEC = 1.01 DN for all Agg. Seq.
  - LGS NEC = 0.74 DN for all Agg. Seq.
- After eliminating data effected by solar stray light using the solar nadir angle, there are about 10,000 samples per day for each detector, HAM side, Agg Seq. and gain stage combination.
- This number of samples is sufficient to determine offset drift for for the daily mean
  - high-gain Agg. Seq. 32 to within a HGS uncertainty = 0.03 DN.
  - MGS and LGS the uncertainty = 0.01 DN.

Conclusions & Recommendations (3 of 3)

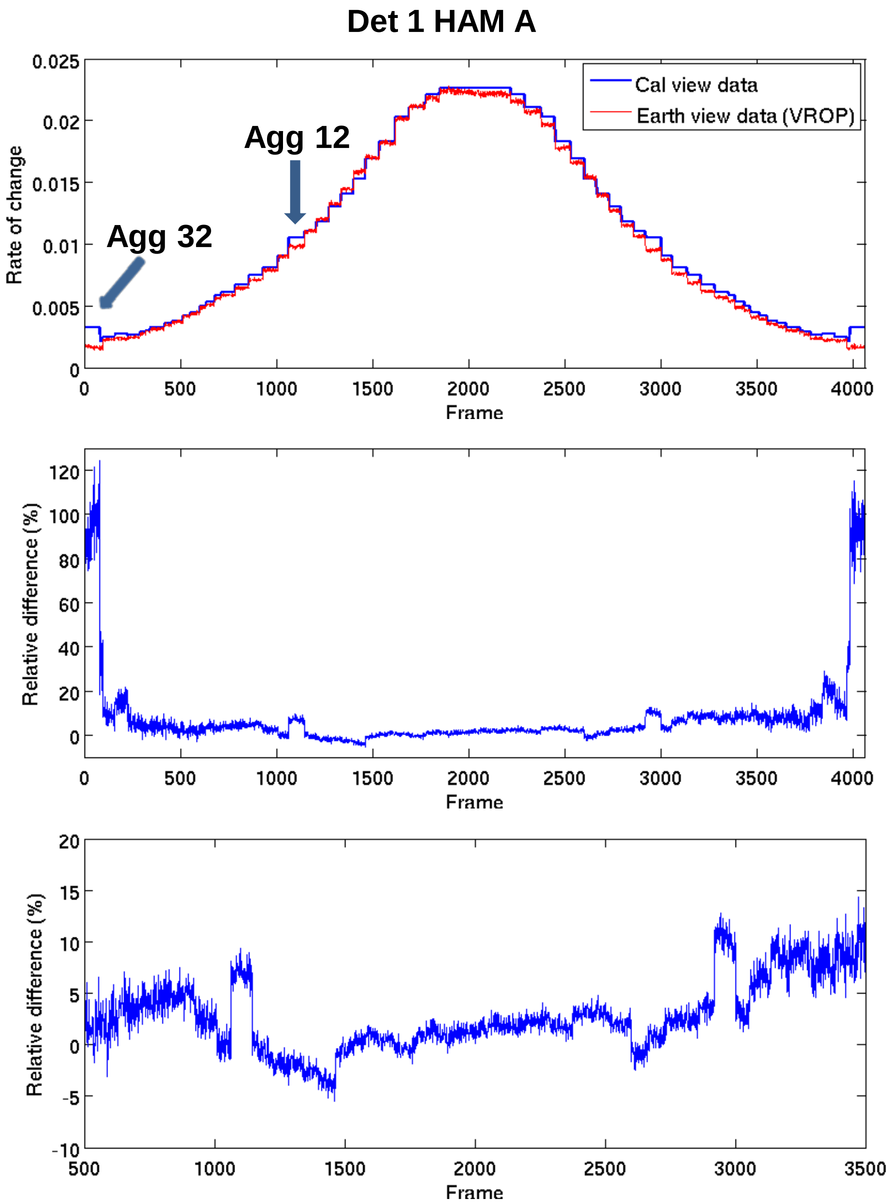
- The method used for outlier mitigation has a large impact on the uncertainty of the daily mean offset as well as the NEC.
  - Modeling showed that Winsorization had the least impact on uncertainty, and this is the process that is recommended here.
  - Trimming and Multilayer Median Trimming were also considered but did not perform as well.
- More important than the method used for outlier mitigation is the limits used.
  - These should not be arbitrarily set but should be set by first determining the probability of outliers that are mostly due to HEP hits.
  - It is recommended that this probability should be determined using either Peirce's criterion and/or Chauvenet's criterion to identify likely outliers.
- There may be sufficient indirect light during a full moon to produce a detectible signal in the daily mean offsets from BB.
  - These events should be studied and if necessary removed from the trending using a threshold lunar phase.
- Outlier processing should also be performed on the daily mean offsets after removing the impact of drift.
  - This may be need to remove anomalous events that occur over a period longer than a few seconds.

Part2—Backup

DNB Cal Data Form

- As with all the other bands on VIIRS, the DNB views the space view (SV), black body (BB) and solar diffuser (SD) calibration (cal) views one per scan. There are two differences regarding the DNB functionality compared with the other bands:
- It has 4 gain stages, HGA, HGB, MGS and LGS, compared with 1 or 2 stages compared with the other bands;
- It has 32 aggregation modes plus 4 diagnostic aggregation modes compared with 3 aggregation modes for the other bands.
- To accommodate these differences, the DNB cal data are acquired differently. The features of the data that are different are:
- There are only 16 samples per scan per gain stage for DNB (compared with 48 for Moderate Resolution bands and 96 for imagery bands).
- The aggregation modes cycle through every 72 scans with 2 scans for each mode (one for each HAM side).
- Aggregation modes are numbered from the center at 1 to the edge at 32
- Aggregation sequences are numbered from the <u>edge</u> at 1 to the center at 32

Rate of Offset Change Comparison

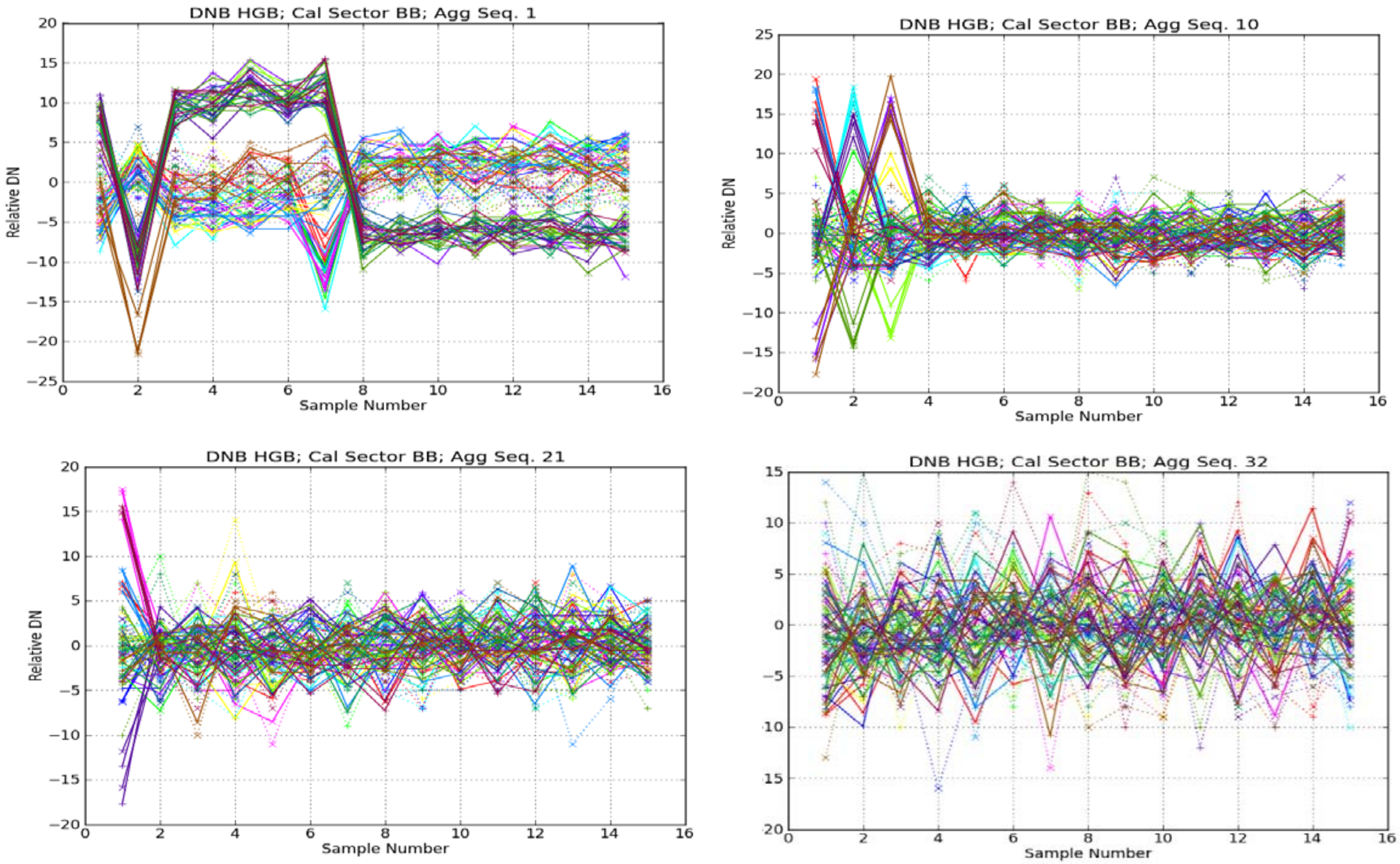


Top: HGS rate of change fitted from 47 new moon days (02/21/2012 and 46 days between 11/13/2012 and 07/04/2016).
Cal view data: follow the RSBAutoCal algorithm to determine DNB dark signal.
Earth view data (VROP): DNB DN0 LUT (HGS)

Middle: relative difference of the fitted change of rate (rate_CalView - rate_EarthView)/rate_EarthView.

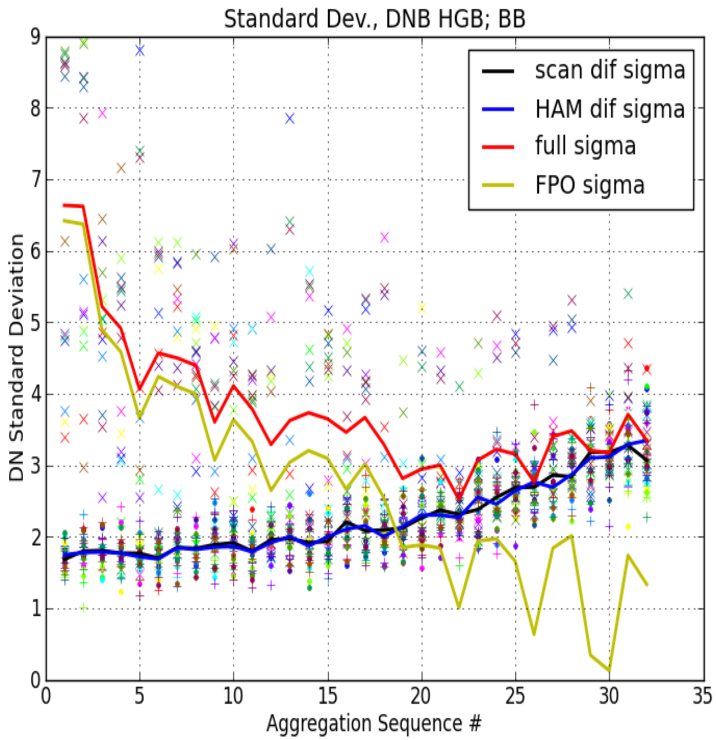
Bottom: zoomed in figure of the middle figure

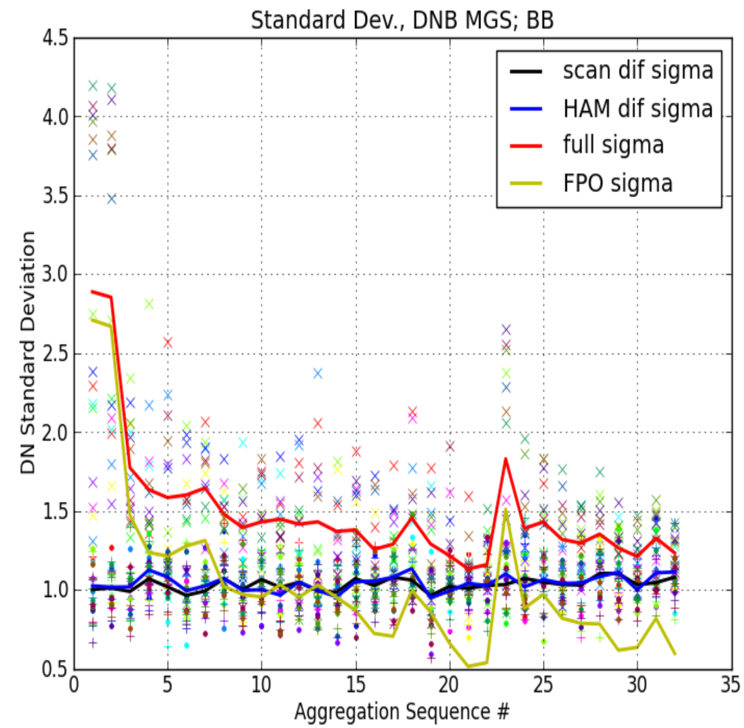
Fixed Pattern Offset Dominates HGB Cal



HGB Fixed Pattern Offsets for 4 aggregations. Solid lines are offsets, dotted lines are differences plotted to show noise level. Both HAM sides are plotted with "+" for side 0 and "x" for side 1. Color indicates detector number.

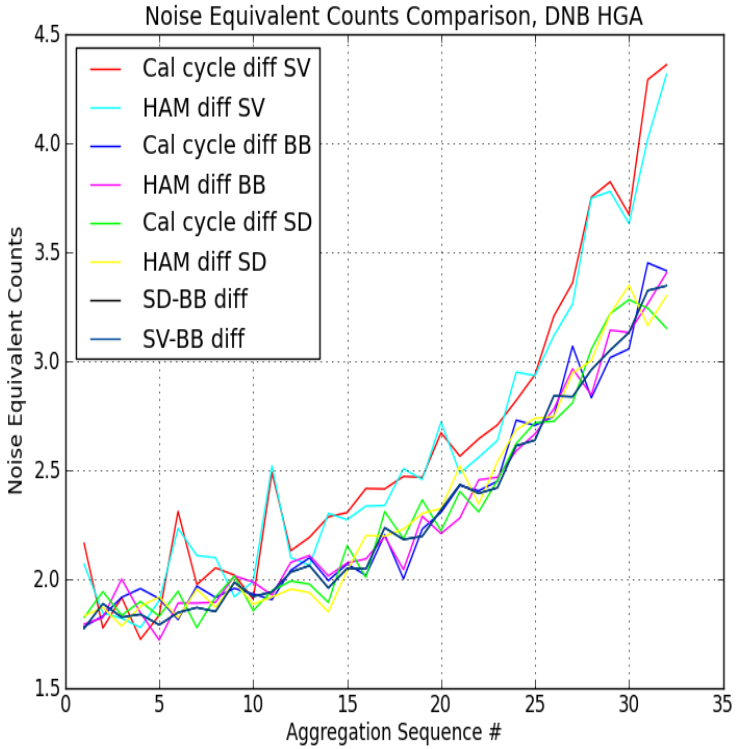
Detector noise and FPO variation compared for HGB

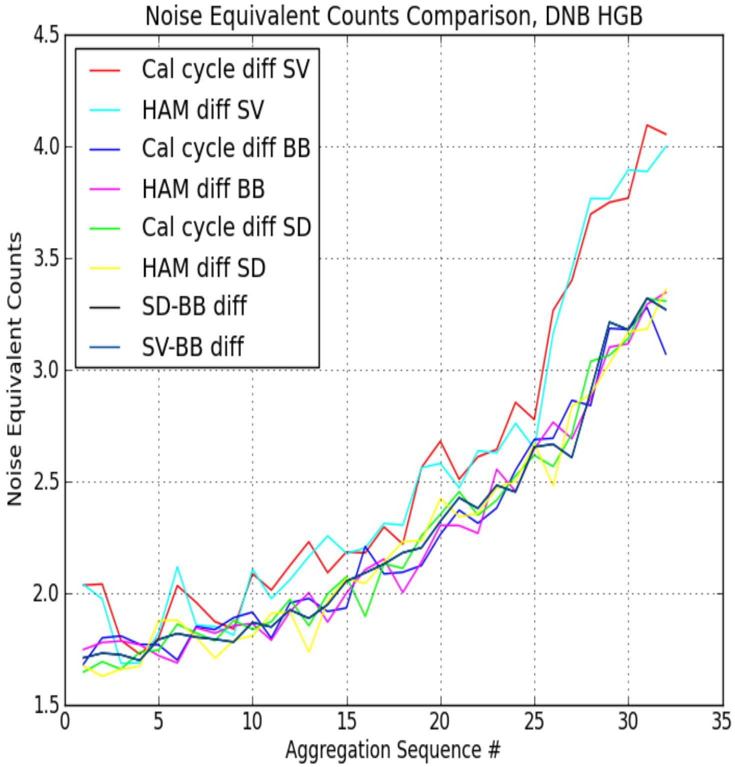




FPO variation for HGB is different from HGA, but it usually dominates for lower Aggregation Sequence Number

HGA & HGB Dark Noise Equivalent Counts (NEC) Computed 8 Ways





- Cal cycle SV difference is large due to airglow variation and long time span
- HAM SV likely has higher NEC due to airglow variation

Part 3—Determining offsets using Earth view—a statistical method using a parametric model with method of moments estimator

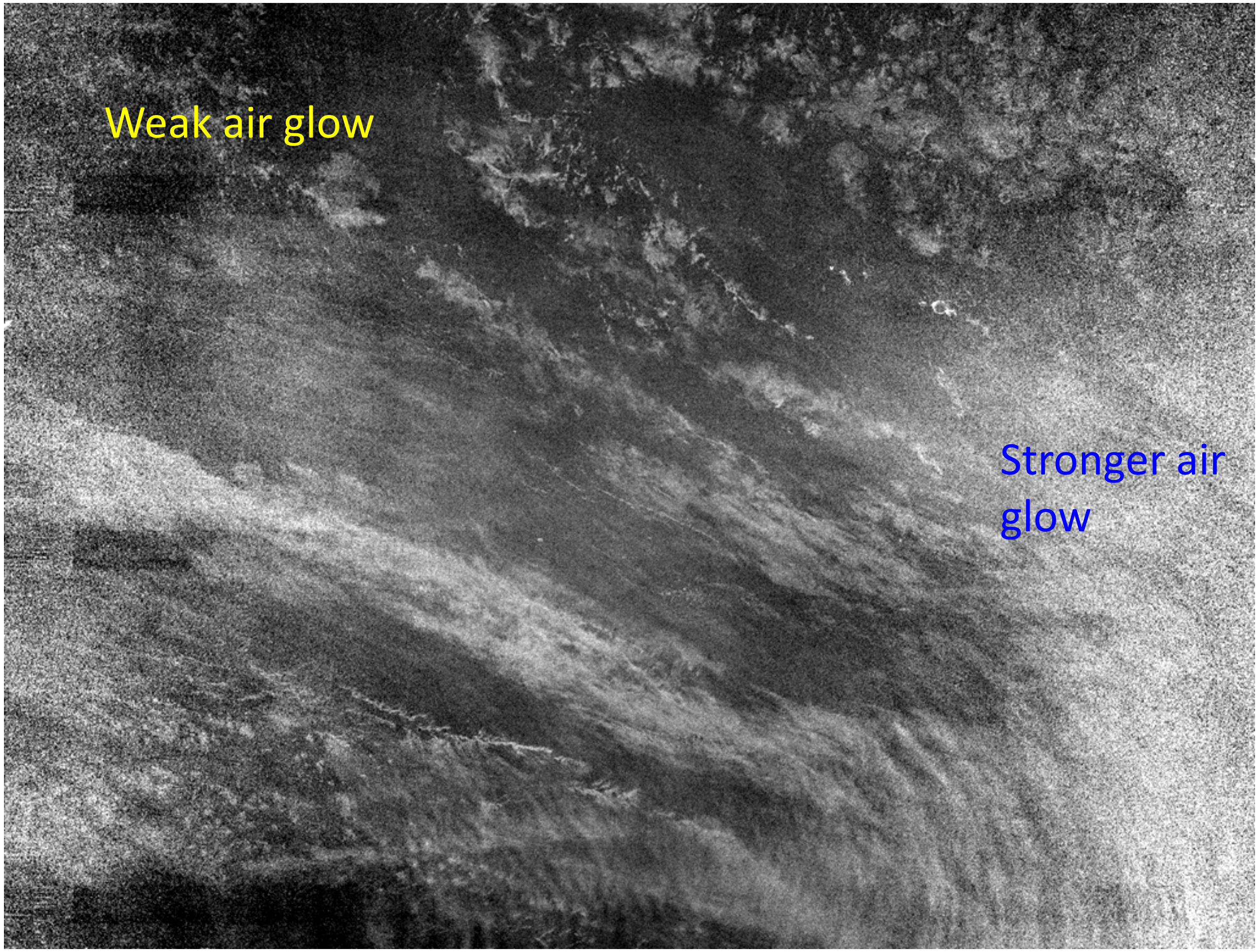
Summary of methodology

- The primary method for determining dark offset is to view the earth at night over the Pacific during a new moon
- but even without any lunar illumination, there is always some detectable light coming from the earth.
- This makes it difficult to use dark earth view data to determine offset tables that are not biased.
- What is proposed here is to use a statistical estimator and a parametric model of the natural illumination to determine the level of natural illumination and therefore correct for it.
- I considered for this are Maximum Likelihood Estimation (MLE) and the Method of Moments Estimation (MME).
- Here only MME will be considered, but it would be worthwhile to investigate MLE.

What data to use

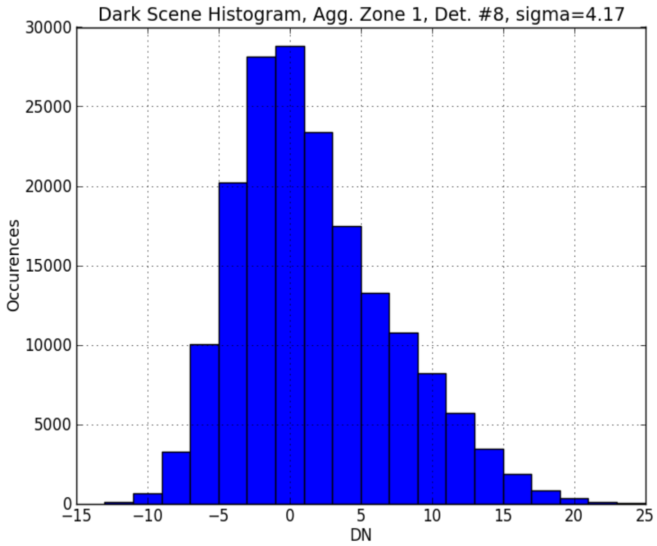
- When determining the MGS and LGS offsets only the data from VROP 702 and 705 are available to be used, but for HGS this is not a restriction.
- The HGS offset data has always been taken over the Pacific Ocean on the day of the new moon, but is it really necessary to restrict the data to only these exact new moon dates?
- The reason for using the Pacific is because there is very few fixed lights, but this is just as true for the Indian Ocean or most of the Atlantic Ocean.
- Also there are places on land such as the Sahara Desert which are similarly deplete of fixed light.
- Because there are databases that provide data on persistent light over the entire earth, recommend that rather than restricting the data collection to certain regions, it is better to use <u>all</u> data and then filter out pixels where there are persistent lights based on the geolocation of the pixels.

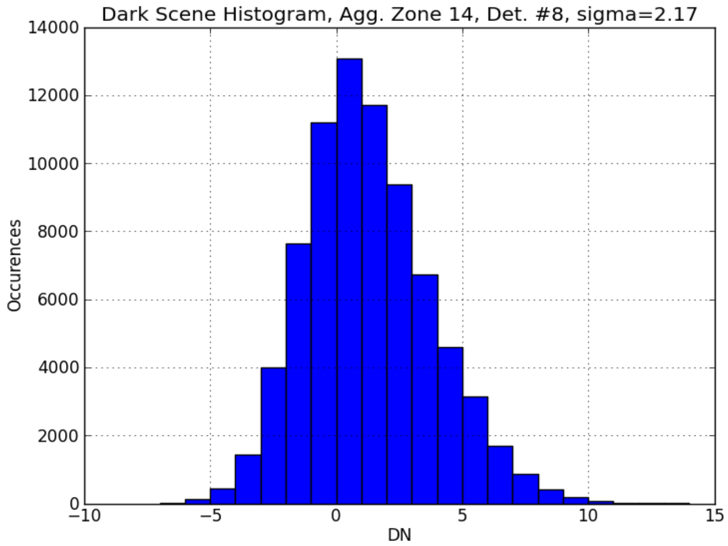
Ocean With no lunar illumination

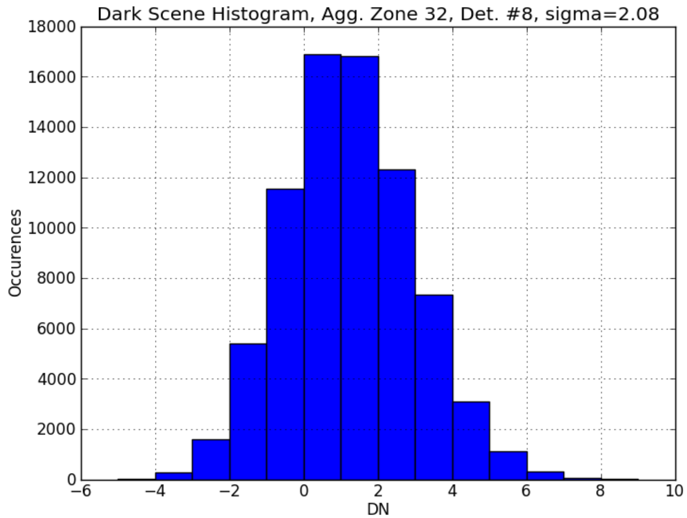


HGS granule taken on 22 Sep 2014 between 11:47:052 and 11:52:46 UTC, during a new moon in the region over the South Pacific Ocean. The plotting range from black to white is from -8.3×10^{-11} to 4.2×10^{-10} W cm^{-2} str^{-1}.

Distributions from new moon DNB scenes







Distribution shape changes with aggregation zone. More skew near nadir.

Math model for scene radiance

Types of illumination:

- Unnatural nighttime illumination (UNI) includes electrical lights, gas flares and other anthropogenic fires.
- Natural Nighttime Illumination (NNI):
  - extraterrestrial nighttime illumination (ETI), including stars, zodiacal light and planets.
  - Airglow, both direct and reflected.
  - natural terrestrial sources other than airglow, including the Aurora Borealis, Aurora Australis, lightning, algae blooms and natural fires.

Remove through filtering

$$L_{all} = \cancel{L_{UNI} + L_{terr}} + L_{ET} + L_{AG} \text{ ,}$$

where: L_{UNI} = radiance of UNI; L_{terr} = radiance from natural terrestrial sources other than airglow; L_{ET} = reflected ETI radiance; L_{AG} = reflected and emitted airglow radiance

$$L_{NNI_n} = E_{ET}(d, t_{sol}, lat) \cdot \rho_n + (c_i \cdot \rho_n(\theta_i) + \sec\theta_i) L_{AG_{emis_n}}$$

$E_{ET}(d, t_{sol}, lat)$ = extraterrestrial irradiance as a function of date, solar time and latitude respectively; and ρ = surface reflectance, and *n* indicates the ensemble index. E_{AG} = downward airglow irradiance; ρ = surface reflectance in units of steradians^{-1}; θ =satellite zenith angle on the earth; and $L_{AG\ emiss}$ = upward total column airglow emission when θ =0

Equate sensor response to scene radiance model

$$DN_{LUT\,j,s,h,k} = DN'_{LUT\,j,s,h,k} - DN_{NNI\,j,s,h,k,m}$$

Where $DN_{LUT\,j,s,h,k}$ is the ideal offset LUT,

$DN'_{LUT\,j,s,h,k}$ is the offset LUT as it is currently produced from

$DN_{NNI\,j,s,h,k,m}$ is the NNI offset that we wish to determine.

indices j, k and m, referring to detector number, aggregation zone number and latitude bin respectively. The index n refers to a random sample of the ensemble, s refers to the sample within the aggregation zone and h refers to HAM side.

$$L'_{n,j,s,h,k,m} + L'_{NNI_{j,s,h,k,m}} = E_{ET_m} \cdot \rho_{n,k,m} + \left(c_{k,m}\,\rho_{n,k,m} + \sec\theta_k\right) L_{AG_{emis}\,n,m} + L'_{noise_{n,j,k}}$$

$$L'_{NNI_{j,s,h,k,m}} = DN_{NNI_{j,s,h,k,m}} / g_{j,k}\ ;$$

$$L'_{n,j,s,h,k,m} = \left(DN_{n,j,s,h,k,m} - DN'_{LUT_{j,s,h,k}}\right) / g_{j,h,k}$$

$$L'_{noise_{n,j,k}} = N_{n,j,k} / g_{j,k}\ ;$$

Where: $g_{j,k}$ =gain; $N_{n,j,k}$ =noise counts; $DN_{NNI\,j,s,h,k,m}$ =

$DN_{n,j,s,h,k,m}$ = measured counts

Ugly Math!

Produce simultaneous equations using 1st, 2nd & 3rd moments (mean, variance & skew)

we make the assumption that $L_{AG_{emis}}$ has a gamma distribution:

$$\langle L_{AG_{emis}{}_{n,m}} \rangle_n = \int_{-\infty}^{\infty} L^{\alpha_m} / [\beta_m{}^{1+\alpha_m} \Gamma(1+\alpha_m)] e^{-L/\beta_m} \, dL = \beta_m \, (1-\alpha_m)$$

Uglier Math!

Mean:

$$\overline{L'_{NNI}}_{k,m} = E_{ET}{}_m \langle \rho_{n.m,k} \rangle_n + \beta_m \, (1-\alpha_m) \cdot \left[c_{k,m} \langle \rho_{n,m,k} \rangle_n + \sec \theta_k \right] - \langle L'_{n,k,m} \rangle_n$$

There are 4 unknowns here: $\overline{L'_{NNI}}_{k,m}$, α_m, β_m and $c_{k,m}$

Variance:

$$\sigma^2 (L'_{n,k,m})_n = \sigma^2 \left[E_{ET}{}_m \cdot \rho_{n,k,m} + \left(c_{k,m} \, \rho_{n,k,m} + \sec \theta_k \right) L_{AG_{emis}{}_{n,m}} + L'_{noise_{n,k}} \right]_n$$

Skewness:

$$\langle \left(L'_{n,k,m} + L'_{NNI_{k,m}} \right)^3 \rangle_n = \langle \left[E_{ET}{}_m \cdot \rho_{n,k,m} + \left(c_{k,m} \, \rho_{n,k,m} + sec \, \theta_k \right) L_{AG_{emis}{}_{n,m}} + L'_{noise_{n,k}} \right]^3 \rangle_n$$

This gives 192 equations and 130 unknowns. Solve using Newton-Raphson method and regression

Conclusions

- Because of the statistical nature of the sensor noise and scene radiance, a statistical estimator seems to be the only way to solve this problem
- More work is needed to develop this methodology
- The math is complicated but it should not be difficult to program in a language such as Python or IDL.
- In addition to providing unbiased offsets, this method would also provide a mathematical model for understanding airglow distribution