**Vera C. Rubin Observatory
Data Management**
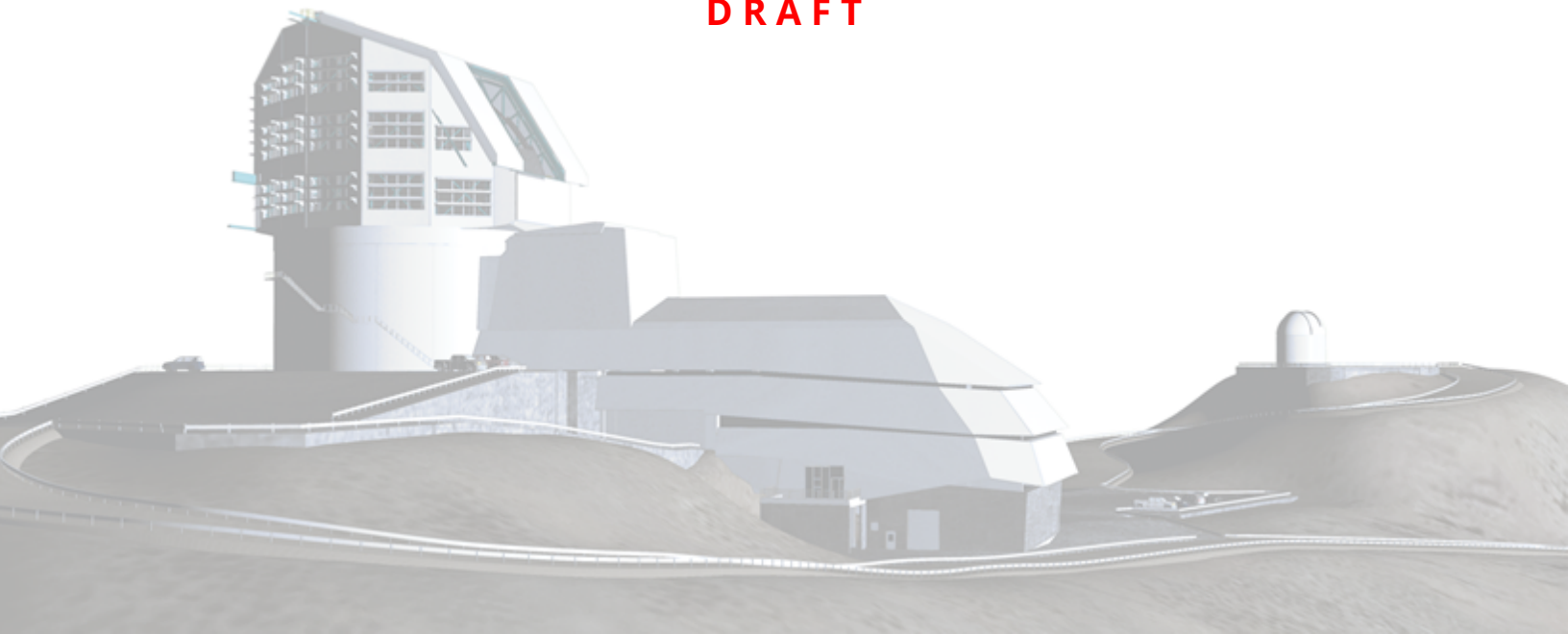
Assessment of Cosmic Ray Rejection for Image Differencing

Eric C. Bellm

DMTN-315

Latest Revision: 2025-05-09

D R A F T

Abstract

We evaluate the contamination rate of DIASources by unmasked cosmic rays (CRs), with implications for the snaps observing mode.

We estimate that after filtering on CR pixel flags and apply a reliability score threshold, fewer than 3% of real ComCam transients would be due to unmasked CRs. We outline several straightforward steps to further improve rejection performance on LSSTCam. Data taken in snaps mode does not appear to be a panacea due to non-CR changes in alignment, PSF, saturation, etc. between the two snaps.

Change Record

Version	Date	Description	Owner name
1	2025-05-08	Unreleased.	Eric Bellm

Document source location: `https://github.com/lsst-dm/dmtn-315`

Contents

Assessment of Cosmic Ray Rejection for Image Differencing

1 Introduction

Cosmic rays (CRs) are a potential source of contamination in image differencing. The baseline plan for Rubin Observatory was to use "snaps"–two consecutive exposures of the same field– to enable rejection of CRs by differencing the snap pair, which would then be coadded and processed as a single exposure. Other surveys have instead utilized morphological algorithms and/or machine learning to identify and mask CRs based on their sharp features.

Given the high cost in observing time of the snaps mode, which requires an additional 2 second read in the middle of the snap pair and costs about 7% of the available survey time, in this technote we evaluate the contamination rate of DIASources by unmasked CRs.

2 Pipelines and Data

We used the DRP processing of ComCam data conducted on DM-49235 using weekly `w_2025_09`. This processing differs slightly from that released in DP1 but the changes are not expected to affect the results presented here. The standard processing pipeline uses a morphological algorithm to identify and interpolate over CRs in the science image. CRs are not expected to be present in the template image due to artifact rejection, which should be effective given the large number of input exposures available in this dataset. After differencing, a machine learning algorithm was applied to cutout triplet images to assesss the probability that the DI-ASource was astrophysical. This "Real/Bogus" algorithm was largely trained on injected point sources.

We selected one DIASource at random from each of the 15,962 detector-visit DIASource catalogs. For each selected DIASource, we loaded a png image of the science, template, and difference image cutouts in the `tasso` labeling interface. Users were asked to label the cutout triplets with one of several real or artifact labels. These included a cosmic ray label as well as a "real transient" label. The latter indicated a point source present in the science image but not in the template. This process yielded 9522 useful labels of unique DIASources.

3   Results

For this analysis, we focused on sources labeled as "real transients" or "cosmic ray". Relatively few trailed sources (13) were present in our labeled sample, and while CRs can affect photometry for variable stars and AGN the scientfic cost of false positives is less significant than for potential transients or non-trailed solar system objects. We rechecked the labels of the transients and CRs and removed a number of incorrect labels. In total, the sample contained 99 real transients and 54 CRs[1].

While at first glance the contimanation by CRs seems high, further filtering can remove many of the CRs. Of the 54 CRs, 47 had `pixelFlags_crCenter` True, in contrast to only 1 transient. It appears that the majority of the CRs producing false positives are on the edges of larger CRs which have been masked by the morphological algorithm. While they have not been properly interpolated, the presence of the CR pixel flag can effectively identify them.

Of the remaining 7 CRs without the relevant pixel flag, three were CRs appearing in the the template and the remainder were genuinely missed. Of these, only 3 had Real/Bogus scores above 0.1.

We therefore estimate the contamination rate of transient DIASources in this ComCam processing by CRs to be 3–7% depending on the Real/Bogus threshold used. Including alerts from real variable sources, the effective contamination rate is reduced by a factor of ten.

CRs are a small minority of false positives in ComCam. Detections around bright stars and low-SNR detections in the bluer bands are orders of magnitude more common in current processing, although Real/Bogus scores can filter the majority of these false positives.

4   Snaps Processing

We investigated whether snaps-mode data would be more effective at removing CRs. Using a dataset of 176 ComCam snap pairs, we measured a median relative displacement of (0.13 ± 0.07) arcseconds, or (0.6 ± 0.4) pixels. Although image warping and photometric calibration were applied to correct for this misalignment, residual structures remain visible in the

---

[1]There were also 912 sources classified as real variables, although we did not re-verify this label group. The raw bogus:real ratio is thus about 9:1.

difference images. These are attributed to uncorrected atmospheric and PSF variations.

These residuals present a major obstacle to using the two-snap mode for reliable cosmic ray rejection. In principle, differencing two closely timed exposures should isolate transient artifacts like cosmic rays, assuming that all static sources align perfectly. In practice, however, the residual misalignments produce spurious features in the difference image that can mimic or obscure actual cosmic ray hits. This contamination undermines the reliability of CR detection using snap differencing.

5   Future Improvements

At present, morphological CR identification is not performed on the difference image. We have identified some cases of CRs on top of real sources for which this approach would be useful. However, in our tests, the CR identification algorithm also tends to identify subtraction artifacts with sharp edges as CRs. Further investigation will be needed to determine if this is a useful approach.

We identified a number of cases where the CR interpolation algorithm failed to mask the entire CR. Further tuning may enable better performance.

The Real/Bogus model used in this analysis was relatively basic. Explicit training on CRs that it assigns high scores to would enable improved rejection of these sources.

A   References

B   Acronyms

Acronym	Description
DM	Data Management
DMTN	DM Technical Note