1 Abstract
Several algorithms have been proposed to identify cosmic ray events in non-destructive
ramps. Regardless of the search method, one eventually faces the question of how to
assess if a candidate event is a real CR or a noise fluctuation. The Generalized Least
Square estimator optimally calculates the ramp parameters and CR value, with their
correlated uncertainties. Using these statistical parameters, together with prior knowledge
on the frequency and spectrum of CR events, Bayes theorem allows us to establish the
correct criterion for deciding on the reality of a detected CR. The method is expanded to
account for the case of grouped frames, where the apparent probability of having 2 CRs
in adjacent groups is boosted by single CRs falling within a group. In this case the GLS
can be adapted to recover the position of the CR within the group and restore the ramp in
its integrity. The Bayesian criterion can be very easily implemented in the pipeline once
the ramps are reconstructed using the Generalized Least Square algorithm. Tuning the
priors for optimal decisions will require regular monitoring and analysis of JWST data.

2 Introduction
Cosmic rays (CRs) are expected to represent a significant source of noise in JWST
detectors; they affect the signal-to-noise of the data (Offenberg et al. 2000, Robberto
2008) and can limit the maximum duration of exposures, possibly preventing reaching
background limited conditions in low flux conditions. The frequency and energy
spectrum of CRs at L2 has been measured by various missions, probing different phases
of the solar cycle. For example, the ESA Standard Radiation Environment Monitor
(SREM) onboard the Herschel spacecraft (Horeau et al. 2011) measured at solar
minimum count rates for protons with E>20MeV of about 5 H/s (Hydrogen
nuclei/second) over a surface of 1.44 cm², corresponding to a flux of about 3 H/s/cm²;
electrons with E > 0.5MeV provide about 2 e/cm²/s. This would correspond to about 100
CR/s on each JWST HgCdTe detector.
Several methods have been proposed to isolate cosmic rays, and outliers in general, in a
non-destructive ramp. Offenberg et al. (2000) looked for the worst outliers relative to the
known readout noise and estimated photon shot noise; Anderson and Gordon (2011) compared three methods: a two-point difference method, a deviation from the fit method, and a y-intercept method. Martel (2012) proposed an Extreme Studentized Deviate (ESD) test. Their results, in general, are

1. The discovery of one or more cosmic ray events in a ramp;
2. The measure of the intensity of the cosmic ray(s);
3. The reconstruction of the ramp purged by the cosmic ray(s).

All these method provide excellent results when the CRs are strong. However, as their intensity decreases approaching the variable noise level of the integrated signal, their accuracy degrades for two basic reasons. First, they poorly measure the intensity of the cosmic ray and the ramp parameters, as they neglect the fact that JWST data are taken in non-destructive mode. When the samples are correlated, one has to write the full covariance matrix and perform a Generalized Least Square fit to derive the slope and intercept of the ramp, with their associated statistical parameters (covariance, $\chi^2$, etc.). Robberto (2014) has shown that once a cosmic ray has been tentatively identified, the Generalized Least Square method can be used to solve points 2 and 3 simultaneously to derive a value for the CR intensity that is consistent with the ramp’s slope and intercept, together with their errors and cross-correlations. Attacking point 1, Robberto (2014) also hinted at the possibility of searching for a CR by iterating the GLS algorithm over each sample/group of a ramp as if each one was affected by a cosmic ray, and looking then for a deep minimum in the set of resulting $\chi^2$ values to find a suitable candidate. In other words, one can entirely rely on the GLS to a) search for CRs, and b) reconstruct the ramp.

The second problem is how to decide if a ramp has been really affected by a CR. The risk, of course, is to accept false positives (Type I errors in Hypothesis Testing theory) and reject false negatives (Type 2 errors). One first uses the basic GLS fitter without CR to find the slope and intercept: this is the Null hypothesis. One then uses the extension of the GLS to the case with a third parameter (CR intensity) to see if there is any improvement, rejecting the Null hypothesis. The problem is that the solution assuming the presence of a CR in the ramp (all solutions assuming a CR!), at any point and regardless of its real occurrence, will always return a better $\chi^2$ value than the solution without a CR. This because the addition of a free parameter in the fitting process, the intensity of a CR, always implies a lower $\chi^2$. How can we be sure that the solution with the lowest $\chi^2$, obtained assuming that a CR is present in the ramp, is really preferable to the solution with only 2 parameters, without a CR?

A simpler theory capable of reasonably explaining the data is preferable to a more complex theory. This is the so-called “Ockam Razor” principle that has played a crucial role in the history of science, see e.g. the dispute between Galileo and the Inquisition on the Keplerian vs. Ptolemaic systems. Simplicity can be preferable on the basis of its esthetical appeal, but modern statistics allows us to put the general problem of “model selection” on firm quantitative ground. In this report we will apply this technique to solve the “CR or not CR” model selection problem for ramps taken in non-destructive
mode, accounting also for the case of frames averaged in groups, typical of JWST.

3 A Bayesian approach to CR evaluation

Once a possible cosmic ray of intensity \( C \) has been found at a time \( t_{CR} \) in a linearized ramp of intercept \( a \) and slope \( b \), given the data \( c_i \) collected at the time \( t_i \), i.e.

\[
c_i = a + b \cdot t_i + \Theta(t_i - t_{CR})
\]

having used the Heaviside function, \( \Theta \), to account for the CR event, the problem we are trying to solve can be expressed as a comparison between the probability that the CR is real vs. the probability that it is not. These two hypotheses can be indicated as \( H_{CR} \) and \( H_{\overline{CR}} \). We must discriminate between these two hypotheses on the basis of the collected data \( c_i \), of the derived model parameters, i.e. either a set \( \Theta_{CR} = \{a_{CR}, b_{CR}, t_{CR}\} \) or a set \( \Theta_{\overline{CR}} = \{a_{\overline{CR}}, b_{\overline{CR}}\} \), and of any prior knowledge on the phenomenon, like e.g. the CR flux and expected energy spectrum, before looking at the collected data.

If we were interested in the problem of finding the best-fit parameters according to either hypothesis, CR or no CR, we would deal with the Bayes theorem in the form

\[
P(\theta|c_i,H,T) = \frac{P(c_i|\theta,H,T) \cdot P(\theta|H,T)}{P(c_i|H,T)}
\]

having indicated with \( T \) our knowledge (either from the theory or past experience) on the CR behavior. The 3 terms at the r.h.s. of Eq. (2) represent:

- \( P(c_i|\theta,H,T) \): the likelihood of obtaining the set of measures given the ramp parameters;
- \( P(\theta|H,T) \): the prior probability of obtaining the parameters on the basis of our previous knowledge about the phenomenon;
- \( P(c_i|H,T) \): the evidence, i.e. the probability of obtaining the set of measures given any previous knowledge about the phenomenon. This term is often ignored as it can be regarded as a normalization constant, after integration of the space of possible parameters \( \theta \).

In practice, we solve this problem (“inference”) using the GLS algorithm, counting on the fact that the posterior probability has a strong peak in correspondence to a set of most probable parameters \( \hat{\theta} \), and departs from these parameters following a multidimensional Gaussian described by the covariance matrix of the parameters, which is also returned by the GLS fitting process. We also assume that the GLS algorithm provides us with one (or more) strong candidates at a certain arrival time in the ramp, allowing us to drop the time (or, more precisely, the interval) of arrival from the list of unknown parameters. We can therefore write for the posterior probability:

\[
P(\hat{\theta}|c_i,H,T) \approx P(c_i|\hat{\theta},H,T) \cdot P(\hat{\theta}|H,T)
\]

We are now interested in a different problem, the “decision” about the hypotheses. In this case, the Bayes theorem must be written as
\[ P(H|c,T) = \frac{P(c|H,T) \cdot P(H|T)}{P(c|T)} \]  

and we immediately see that the evidence, the term at the denominator of Eq. (2) that was ignored in the previous problem of finding the parameters, plays here a key role. We are now dealing with the ratio:

\[ \frac{P(H_{CR}|c,T)}{P(H_{CR}|c,T)} = \frac{P(c|H_{CR},T) \cdot P(H_{CR}|T)}{P(c|H_{CR},T) \cdot P(H_{CR}|T)} \]

\[ = K \cdot \frac{P(H_{CR}|T)}{P(H_{CR}|T)} \]

where we have indicated with the letter \( K \) the “Bayes factor”, the ratio of the evidences. The last term is the ratio between the priors on our two models, i.e. a measure our prior expectation of having or not a CR in the ramp.

As stated above, the evidence is the normalization factor in Eq. (2), therefore:

\[ K = \frac{P(c|H_{CR},T)}{P(c|H_{CR},T)} = \int P(c|\theta_{CR},H_{CR},T)P(\theta_{CR}|H_{CR},T)d\theta_{CR} \]

We see now that the evidence can be expressed as the integral of the product of likelihood, \( P(c|\theta_{CR},H_{CR},T) \) times the prior on the parameters, \( P(\theta_{CR}|H_{CR},T) \), integrated over the entire space of the parameters. When we have used the GLS fitting to derive the parameters \( \hat{\theta} \) and their covariance matrix, we have implicitly assumed that the prior on the parameters is uniform, so that this product has a strong peak in coincidence of the maximum of the likelihood; the solution is then concentrated in a small region of the parameter space. In the case of one dimensional distribution, the integral can then be approximated by the integral of a Gaussian, which is proportional to the peak height, \( P(c|\hat{\theta},H,T)P(\hat{\theta}|H,T) \), times the Gaussian width \( \sigma \). In the case of a multivariate normal distribution, the integral can be approximated by the product of the Gaussian peak value by the subtended multi-dimensional volume, i.e.

\[ K = \int P(c|\theta_{CR},H_{CR},T)P(\theta_{CR}|H_{CR},T)d\theta_{CR} \]

\[ = \frac{P(c|\hat{\theta}_{CR},H_{CR},T)P(\hat{\theta}_{CR}|H_{CR},T) \cdot \sigma^3}{P(c|\hat{\theta}_{CR},H_{CR},T)P(\hat{\theta}_{CR}|H_{CR},T) \cdot \sigma^2} \]

Check with the JWST SOCCER Database at: https://soccer.stsci.edu
To verify that this is the current version.
having indicated with the Deltas the multi-dimensional volumes for the case with 3 and 2 parameters (caveat: 3 and 2 are not signs of exponent here). Note that a highly non-uniform prior may bias the integral, but we will assume that this is not the case (see Sec.4).

The first term, \( P(c|\hat{\theta}, H, T) \), is the best fit likelihood, related to the \( \chi^2 \) of the GLS fit by the relation

\[
P(c|\hat{\theta}, H, T) \propto \exp\left(-\frac{\chi^2_{\text{GLS}}}{2}\right); \tag{8}
\]

displaying the term that gives more weight to the best fit, the one with smaller residuals. The other two terms

\[
P(\hat{\theta}_{\text{CR}}|H_{\text{CR}}, T) \cdot \sigma_{\hat{\theta}_{\text{CR}}}^3, \tag{9}
\]

and

\[
P(\hat{\theta}_{\text{CR}}|H_{\text{CR}}, T) \cdot \sigma_{\hat{\theta}_{\text{CR}}}^2 \tag{10}
\]

represent the Occam factors. They penalize the model with more parameters because the prior range over which the parameters are specified are generally much larger than the range of acceptable likelihoods. In our case, this means that, having found a CR at a certain energy and uncertainty (the posterior), we weight our success against our a priori expectation of having a CR of that energy in the ramp. Since the integrals in Eq. (6) are carried out over all parameters, we should also include the Occam terms relative to the slope and intercept; in practice, we can safely assume that they are independent of the occurrence of a CR, so their integrals cancel out once the ratio of the two models is taken.

The Delta terms, i.e. the multi-dimensional volumes subtended by the multi-dimensional Gaussian, can be immediately estimated from the covariance matrix of the coefficients \( \Sigma_{\hat{\theta}} \), or Hessian, i.e. the \( 2 \times 2 \) or \( 3 \times 3 \) error matrix returned by the GLS fitting procedure. It is, dropping all indexes of the sigma:

\[
\sigma = \det\left(\frac{1}{2} \left(\Sigma_{\hat{\theta}} \right) \right) \tag{11}
\]

This term accounts for the errors in the parameters and their correlations.

To summarize, we can write Eq.(6) as:

\[
K = \frac{\exp\left(-\frac{\chi^2_{\text{GLS}}}{2}\right) \cdot P(\hat{\theta}_{\text{CR}}|H_{\text{CR}}, T) \cdot \det\left(\frac{1}{2} \left(\Sigma_{\hat{\theta}_{\text{CR}}} \right) \right)}{\exp\left(-\frac{\chi^2_{\text{GLS}}}{2}\right) \cdot P(\hat{\theta}_{\text{CR}}|H_{\text{CR}}, T) \cdot \det\left(\frac{1}{2} \left(\Sigma_{\hat{\theta}_{\text{CR}}} \right) \right)} \tag{12}
\]

In conclusion, Bayesian model selection can be regarded as an appendix to the problem of finding the most likely parameters of each model. All what is needed is to multiply the likelihood (an exponential of the \( \chi^2 \) returned by the GLS fit) by our prior expectation on the probability of the parameters (i.e. our understanding of the CR intensity spectrum) times a function of the Hessian matrix (also returned by the GLS fit), assuming the

Check with the JWST SOCCER Database at: https://soccer.stsci.edu

To verify that this is the current version.
Gaussian approximation is adequate; the last term is our second prior, accounting for the a-priori probability that a CR has actually impacted the signal ramp. In the next section we will show how to derive the priors.

4 Determining the priors

Assuming the priors depend only on the cosmic rays and not on the ramp slope and intercept, we need to quantify the a priori probability of having that CR in the ramp and the probability of not having any CR in the ramp. The prior probability of having that CR in the ramp is the product of two terms; the probability $P(H_{CR}|T)$ of having $k$ CRs in the ramp, which will depend on the rate of arrival $\lambda$ (events/s), and the probability $P(\hat{\theta}_{CR}|H_{CR} \cdot T)$ that a CR arrives with the estimated energy; we can write

$$P(H_{CR}|T) = P(k|f(\lambda))$$

$$P(\hat{\theta}_{CR}|H_{CR} \cdot T) = P(E|g(E))$$

(13)

to indicate that our theoretical knowledge is encapsulated in two functions, $f(\lambda)$ for the cosmic flux and $g(E)$ for the energy spectrum. More precisely, this second equation must account for the amount of charge, or DN, released by a CR in a pixel, which depends not only on the cosmic energy spectrum, but also on the geometry and material of the spacecraft (shielding) and on the physics of the detector; plus other factors, as CRs can also upset the readout electronics appearing e.g. as negative discontinuities in the count rate. Also, in principle, the two functions should be correlated: it is known that the number of high energy CRs decreases at solar maximum as the heliosphere expands, shielding the inner solar system from the so-called galactic CRs; by adjusting our expectations to the date of the experiment and to the occasional bursts of Sun activity we can treat the two functions as independent, so we shall proceed under this assumption.

Let us consider first the first prior, the rate of CRs. If we make the usual assumption that CRs follow Poisson statistics, the probability of having $k$ CRs in a ramp will be given by

$$P(k|f(\lambda)) = e^{-\lambda \tau_{ng}} \left( \frac{\lambda \tau_{ng}}{k!} \right)^k$$

(14)

where $\lambda$ is the rate of CR events per second, $\tau$ is the readout time per group and $n_g$ is the number of groups in the ramp; $\tau n_g$ is thus the duration of the ramp, from reset to the last read. If there is a single CR, $k=1$, it is

$$P(k=1|f(\lambda)) = e^{-\lambda \tau_{ng}} \lambda \tau_{ng}$$

(15)

To give a value, a flux of 1 CR/cm$^2$/s corresponds to $\lambda = 3.24 \times 10^{-6}$ CR/pixel/s for 18$\mu$m pixels. If $\tau = 10.7$ s and $n_g = 100$, corresponding to about 1,000 s integration time, we have $P(k=1) = 3.45 \times 10^{-3}$, or 1 CR event every 15$\times$15 pixels. The probability of having more than one CR in a ramp is much smaller: $P(k=2) = 1.20 \times 10^{-7}$, i.e. about one ramp with two CRs every two integrations 1,000s long. The real flux may be higher and CRs may affect more than one pixel, so we will pay attention to the probability of 2 CRs, but it seems safe to ignore the probability of having 3 or more CRs in a ramp.

Check with the JWST SOCCER Database at: https://soccer.stsci.edu
To verify that this is the current version.
If we include 2 CRs, we must account for the (very small) probability that they both fall within the same group (this case will be more important in the next section, where we discuss grouped frames). A pair in the same group is practically indistinguishable from a single CR; it can readily accounted for by adding a second term to Eq.(15):

\[
P(k = 2|f(\lambda)) \frac{1}{n_g} e^{-\lambda \tau_n} \left(\frac{\lambda \tau_n}{2n_g}\right)^2
\]

(16)

where the division by \(n_g\) selects the single group out of their total number. This increases the “effective” probability of having 1 CR, so that adding the two terms we have:

\[
P(k = 1|f(\lambda)) = e^{-\lambda \tau_n} \lambda \tau_n \left(1 + \frac{\lambda \tau_n}{2n_g}\right)
\]

(17)

for the numerator. For the denominator, we need to subtract from 1 the probability of having 1 or 2 CRs, i.e.

\[
P(k = 0|f(\lambda)) = 1 - P(k = 1|f(\lambda)) - P(k = 2|f(\lambda))
\]

\[
= 1 - e^{-\lambda \tau_n} \lambda \tau_n \left(1 + \frac{\lambda \tau_n}{2n_g}\right)
\]

(18)

Finally, if we find two cosmic rays in different groups, the probability is

\[
P(k = 2|f(\lambda)) \frac{n_g - 1}{n_g} = e^{-\lambda \tau_n} \left(\frac{\lambda \tau_n}{2n_g}\right)^2 \frac{(n_g - 1)}{n_g}
\]

(19)

In what concerns the other prior, \(P(E|g(E))\), or better \(P(DN|g(DN))\) as we deal with detected counts, it only appears in the numerator in Eq.(12). As it is difficult to predict a priori the spectrum of detected CRs, including the negative ones, it is better to rely on experimental data, i.e. on the huge amount of images collected by the instrument itself if a suitable monitoring and analysis program is set in place. One could then be able to build plots similar to those presented in Figure 1, relative to the four CCDs of WFPC2. According to the WFPC2 instrument handbook “[…] A good approximation to the cumulative distribution of events as a function of total signal is given by a Weibull function with exponent 0.25. This function has the form:

\[
N(S) = N_0 \exp[-\lambda (S^{1/4} - S_0^{1/4})]
\]

where \(N\) is the total number of events which generate a total signal larger than \(S\). The best fit to the observed events gives \(N_0 = 1.4\) events chip\(^{-1}\)s\(^{-1}\), \(S_0 = 700\) electrons, and \(\lambda = 0.57\). The fit fails below \(S_0\) and should not be extrapolated to low-signal events. The rate of events with total signal below 700 electrons is 0.4 events chip\(^{-1}\) s\(^{-1}\) (i.e. total events per CCD per second is \(N_0 + 0.4 = 1.8\)).”

The low energy tail, nicely flat, reassures us that the assumption of uniform prior is appropriate at low counts. To calculate the prior, while it would be ideal to have analytical approximation for the observed flux of CRs, a tabular form may be needed. One will have to pay special attention to the evaluation of the low energy tail of CRs in

Check with the JWST SOCCER Database at: [https://soccer.stsci.edu](https://soccer.stsci.edu)
To verify that this is the current version.
order to establish the correct prior. This may require processing the dark current images with aggressive CR removal strategies; comparing the $\chi^2$ distributions running the GLS fitter with and without CRs one can identify outliers.

Once the distribution $g(DN)$ is known, the prior is given by the ratio between the area subtended by the CR event, measured in counts as $\hat{D}N \pm \sigma_{\hat{D}N}$, and the total are subtended by the curve, i.e.

$$P(E|g(E)) = \frac{\hat{D}N \cdot \sigma_{\hat{D}N}}{\int g(DN)dDN}$$

(20)

The panel at the bottom-right of Figure 1 shows an example.

![Figure 1 Distribution of total energy of cosmic ray events in WFPC2. The very small number of events and the relative flatness of the distribution below about 30 DN seems to indicate that these events are rare in WFPC2. The red area shows the measured location of a CR event; the ratio of the red/total area represents the term to be used in the prior.](image)

In summary, the priors in the case of 1 CR, can be written as

$$\frac{P(H_{CR}|T) \cdot P(\hat{\theta}_{CR}|H_{CR},T)}{P(H_{CR}|T)} = e^{-\lambda \tau n_g} \lambda \tau n_g \left(1 + \frac{\lambda \tau n_g}{2n_g}\right) \frac{\hat{D}N \cdot \sigma_{\hat{D}N}}{\int g(DN)dDN}$$

(21)

which contains the ramp parameters $\tau$ and $n_g$, well known, the integral of the energy distribution $\int g(DN)dDN$, also a well known number once a monitoring program has been set in place, and the intensity and error on the CR, $\hat{D}N$ and $\sigma_{\hat{D}N}$, returned by the GLS procedure. The case of 2 CRs is immediate:
In conclusion, we underline that the calculation of the priors is also straightforward and does not add significant computation time to basic ramp processing with the GLS method.

5 Ramps with average frames.
If a ramp is made of averaged frames, the likelihoods in Eq.(12) are expressed in the same way as exponential functions of the $\chi^2$. The terms with the Hessian determinant also are the same. The priors in the presence of one or two CRs (numerator), however, will be different.

Assuming that we average groups of $\bar{n}$ consecutive frames (say $\bar{n} = 2$) and skip sets of $\bar{n}$ consecutive frames (say $\bar{n} = 18$), to adopt the NIRCam DEEP2 mode, and that we repeat this sampling to build a ramp of $n_g$ groups, if the CR falls in a gap the ramp will show the typical discontinuity reflecting the fact that it has been hit by a CR; but if the CR falls within averaged frames, the ramp will show two discontinuities reflecting the fact that we are averaging frames without and with the CR signal. This “double jump” will be (should be!) identified as adjacent CRs, a pretty uncommon event in ramps of ungrouped frames: what was a rare occurrence becomes here much more frequent.

It is possible, on the other hand, to analyze the “double jump” as a single CR by slightly modifying the GLS estimator. Let us remind that the main ingredients to the GLS fitter, in the case of a single CR, are the vector of the data:

$$c_i = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

(23)

the vector of unknown coefficients, slope, intercept and CR intensity:

$$\alpha = \begin{pmatrix} a \\ b \\ CR \end{pmatrix}$$

(24)

and the matrix of parameters, reproducing Eq.(1)

$$A = \begin{pmatrix} 1 & t_1 & 0 \\ 1 & t_2 & 0 \\ \vdots & \vdots & \vdots \\ 1 & t_{CR} & 1 \\ \vdots & \vdots & \vdots \\ 1 & t_n & 1 \end{pmatrix}$$

(25)

together with the covariance matrix of the counts. If we have 2 CRs in a ramp falling in
adjacent frames, we can solve using instead of Eq.(24):

$$\alpha = \begin{pmatrix} a \\ b \\ CR_1 \\ CR_2 \end{pmatrix}$$

(26)

and instead of Eq.(25):

$$A = \begin{pmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{n} & 1 & 1 \end{pmatrix}$$

(27)

On the other hand, we can use our knowledge of the fact that the CR arrived within a group and that its presence has affected some, but not all, of the values averaged in that group. For example, if a group is an average of 4 frames, a CR falling between frame 1 and 2 will affect frames 2, 3 and 4; a CR falling between group 2 and 3 will affect frames 3 and 4; and a CR falling between group 3 and 4 will affect group 4. These 3 cases are described by the following set of $A$ matrices:

$$A = \begin{pmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ 1 & t_{CR, 3/4} & 3/4 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & 1 & 1 \end{pmatrix}$$

$A = \begin{pmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ 1 & t_{CR, 2/4} & 2/4 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & 1 & 1 \end{pmatrix}$

$A = \begin{pmatrix} 1 & t_1 & 0 & 0 \\ 1 & t_2 & 0 & 0 \\ 1 & t_{CR, 1/4} & 1/4 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & 1 & 1 \end{pmatrix}$

(28)

that are paired with the unknown coefficients in Eq.(24). In other words, when a double adjacent CR appears in $\bar{n}$ averaged frames, one has to evaluate an extra set of $\bar{n}−1$ GLS solutions to find out where the CR occurred, and the value of the 3 parameters. Having solved for 1 CR, regardless if it falls in grouped frames or in skipped frames, allows us to use the theory developed in the previous section.

6 Bayes factor: CR or not CR?
The Bayes factor $K$ provides us with a comparison between the probability of producing a certain set of data given two different hypothesis. Interpreting its value is a topic discussed in the statistics literature. If $K = 1$ the two probabilities are equal, but as the value
of $K$ increases the probability of the numerator (in our case the presence of a CR) becomes higher. As the evidence provided by the data becomes stronger, one is tempted to accept the hypothesis, but to really accept the hypothesis it is necessary to account for our expectations. With reference to Equation (5), the a-priori probability of the models must be included to compare not the probability of getting the data (the $K$ factor) but the probability of the hypothesis.

If we assume that the case of two CRs in a ramp is truly exceptional, so that there are only two possible models, without and with CR, we can derive a simple formula for the posterior probabilities. When

$$P(H_{CR}|c_i,T) + P(H_{\overline{CR}}|c_i,T) = 1$$

(29)

From the Bayes theorem and Equation (5), it is

$$P(H_{\overline{CR}}|c_i,T) = \frac{P(c_i|H_{\overline{CR}},T)P(H_{\overline{CR}}|T)}{P(c_i|T)}$$

$$= \frac{P(c_i|H_{\overline{CR}},T)P(H_{\overline{CR}}|T)}{K \cdot P(c_i|T)}$$

(30)

and using again the Bayes theorem to rewrite the first term on the right hand side:

$$P(c_i|H_{CR},T) = \frac{P(H_{CR}|c_i,T)P(c_i|T)}{P(H_{CR}|T)}$$

(31)

we have

$$P(H_{CR}|c_i,T) = 1 - P(H_{\overline{CR}}|c_i,T)$$

$$= 1 - \left[ \frac{1}{K \cdot P(H_{CR}|T)} \right] P(H_{CR}|c_i,T)$$

(32)

and solving for $P(H_{CR}|T)$ we get the probability for the Null hypothesis of no-CR:

$$P(H_{CR}|c_i,T) = \frac{1}{1 + \left[ \frac{1}{K \cdot P(H_{CR}|T)} \right]}$$

(33)

If the probability is very low, we reject the null hypothesis and conclude there is a CR; viceversa, if the probability is close to 1, we accept the hypothesis and assume the parameters for a ramp without CR. If the probability is intermediate, we cannot decide:

Check with the JWST SOCCER Database at: [https://soccer.stsci.edu](https://soccer.stsci.edu)
To verify that this is the current version.
both options are viable. To maintain statistical rigor, one can then decide to throw a random number with uniform distribution between 0 and 1 to pick one of the two cases.

Before concluding, it should also be mentioned that taking a decision we have assumed implicitly that the hypothesis on the models we have used are correct. In fact, for true “decision-making”, an extra Bayes factor should be added to account for the possibility that our assumptions are wrong. This extension is critical in many disciplines, starting with economics. In our case, we largely deal with the laws of physics, more reliable than the stock market; still, at solar maximum we should be more uncertain about the expected CR flux because of solar storms. The risk of using wrong assumptions supports the necessity of constant monitoring of the CR flux, both rate and energy, for the entire duration of the mission to make sure that the priors are sufficiently reliable.

7 Example
To illustrate the application of the method, let us consider a few basic examples. Figure 3 shows a ramp of 16 frames obtained with WFC3-IR. The GLS algorithm has been used to find a candidate CR between frames 7 and 8; it falls between two frames that have nearly identical values, with an intensity of -42DN (negative values are usually considered to be the result of CR interacting with the readout electronics, see Offenberg et al. 2000). If we fit the full ramp with the GLS algorithm assuming that there is no CR, we obtain \( \chi^2_0 = 18.3 \); if we fit assuming that the CR is present, we obtain \( \chi^2_1 = 10.2 \), a substantial improvement. The ratio of the likelihoods is \( \exp\left(-\chi^2_1 + \chi^2_0 \right) / 2 \right) = 56.3 \) i.e. we strongly suspect the presence of the CR. However, if we calculate the ratio of the \( P(k|f(\lambda)) \) factors for ramps with 1 CR and no CR, using the same value for \( \lambda \) used in Section 3, \( n_g = 14 \) (the first and second frame are nearly coincident in time) and the total duration of 1405s, we obtain 1/220: this is enough to conclude that the occurrence of a cosmic ray is so uncommon that we must conclude that it is unlikely that there is a CR, regardless on the improvement in fit accuracy.
Figure 2 A randomly selected ramp from a WFC3-IR exposure showing the possible presence of a CR. Our statistical analysis allows us to reject the hypothesis that a CR is present.

Let us now consider a similar case, with a CR dropping -60DN between frames 10 and 11 (Fig. 3). In this case it is $\chi_0^2 = 44.7$ and $\chi_1^2 = 27.5$; the corresponding ratio of the likelihoods is $\exp\left(\frac{-\chi_1^2 + \chi_0^2}{2}\right) = 5,405$ that completely offsets the $1/220$ factor. At this point, the second term has to be taken into account, i.e. the probability of having a CR releasing $-60 \pm 15$ DN (the values are shown in the figure). This requires the knowledge of the expected spectrum, the prior. If the probability is, say, 1/100, then we have a chance 1/22,000 of NOT having that CR, which offsets the likelihood: also in this case, therefore, we would conclude that there is no CR: the rarity of such an event is large enough to overcome the substantial improvement in fit accuracy.

Figure 3 Similar to Figure 2 for another randomly selected ramp from a WFC3-IR exposure showing the possible presence of a CR.

8 Conclusions
We have shown that by pairing Bayes theorem with the GLS estimator, it is possible to derive a general criterion for acceptance or rejection of one or more CRs in non-destructive ramps. The criterion assumes prior knowledge on the rate and energy.

Check with the JWST SOCCER Database at: https://soccer.stsci.edu
To verify that this is the current version.
spectrum of the events, and can be immediately implemented in a GLS algorithm. We also show how to search for CRs when the ramps are made of grouped frames. In this case, the apparent probability of having two CRs in adjacent groups, a rather rare event, appears much higher than single CRs falling within a group affecting two consecutive reads. The GLS can be adapted to recover the position of the CR within the group and restore the ramp in its integrity.

Implementing this approach will be straightforward once the ramps are processed through a GLS fit algorithm. However, for the highest accuracy it will be important to fine tune Bayes’ priors, by implementing a standard procedure to monitor the frequency and spectrum of the CRs falling on each JWST detector using in-flight data.

9 References
Robberto, M., 2014, “A generalized least square algorithm to process infrared data taken in non-destructive readout mode”, SPIE 9143E, 3ZR