
Score-based models for $1/f$ correlated noise correction in James Webb Space Telescope spectral data

Salma Salhi^{1,2,3,4} Alexandre Adam^{1,2,3} Loïc Albert^{1,4} René Doyon^{1,4}
Laurence Perreault-Levasseur^{1,2,3,5,6,7}

¹Université de Montréal, ²Mila, ³Ciela, ⁴IREx Institute, ⁵CCA, Flatiron Institute,
⁶Perimeter Institute, ⁷Trottier Space Institute
{salma.salhi,alexandre.adam,loic.albert,rene.doyon,
laurence.perreault.levasseur}@umontreal.ca

Abstract

The expected atmospheric signal when performing exoplanet transit spectroscopy, especially for terrestrial planets, is measured in a few tens of parts per million. The technique is thus very sensitive to various sources of noise. This is particularly true when using the Single Object Slitless Spectroscopy (SOSS) mode on the NIRISS instrument aboard the JWST, given the wide spectral traces of its images. Current methods to deal with $1/f$ (correlated) noise leave residuals that are almost double that of the expected readout noise. Here, we explore the use of Score-Based Models (SBM) to learn the distribution of noise in dark SOSS images, which we then use as a likelihood model in the Score-based Likelihood Characterization (SLIC) framework to produce posterior samples of the underlying (noiseless) spectral traces. We aim to apply this method to time-series spectroscopic observations, potentially reducing our error to the photon noise limit. This could substantially improve our signal-to-noise by up to a factor of two for some spectral regions and thus enable higher precision transit spectroscopy.

1 Introduction

One of the biggest obstacles to the analysis of astrophysical images is the simultaneous presence of many different sources of noise associated with infrared detectors used in a space environment. These include the readout noise, dark current, epoxy voids, cosmic ray hits, and $1/f$ noise. The latter-most source is a correlated noise structure appearing in images as stripes of varying amplitudes, along the fast readout direction. It affects all near-infrared Teledyne detectors used onboard JWST and is caused by the $1/f$ variability of the reference voltage of each detector's readout electronics [1]. This noise can be difficult to calibrate out, and residuals can creep into calibration reference images. This type of instrumental noise is especially dominant at low flux levels typical of spectroscopic applications with moderate spectral dispersion.

Arguably the current most pressing impact of $1/f$ noise is within the field of exoplanet transit spectroscopy. The Near-Infrared and Slitless Spectrograph (NIRISS)[2] aboard the JWST has an observing mode, Single-Object Slitless Spectroscopy (SOSS)[3] that is optimized for time-series observations of transiting exoplanets. The spatially defocused spectral traces of SOSS hide a larger number of pixels whose $1/f$ correction cannot easily be performed, and, in particular, uncorrected $1/f$ noise residuals in spectrographic images add a source of noise equivalent to doubling the readout noise [3]. Even after correcting for the $1/f$ noise using current methods, the derived precision of the resulting spectral trace can be as much as 2.2 times lower than that of the expected precision [4] at the longer wavelength limit of $2.8 \mu\text{m}$ [2].

The non-Gaussian nature of this noise structure renders it difficult to correct for using current data reduction methods. The typical reduction involves the subtraction of a constant value from each

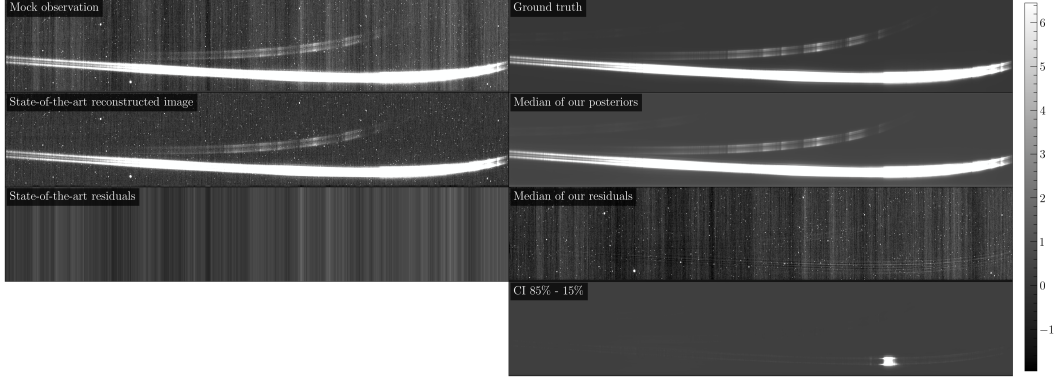


Figure 1: (First row) The mock observation (left) constructed by overlaying the simulated spectral trace (right) and real noise. (Second row) The reconstructed image after applying the conventional $1/f$ correction method (left) compared with the median of 100 posterior samples using our method (right). (Third row) The residuals of observation reconstructed using the state-of-the-art method (left) and the median of the residuals obtained with our method (right). (Fourth row) The 85% – 15% confidence interval on our 100 posterior samples.

individual detector column as shown in the state-of-the-art reconstructed spectral trace in Figure 1, which is done after having subtracted the time-series median frame and masked the spectral traces where the photon noise residuals are dominant. However, this approach leaves residuals that are likely from higher-frequency $1/f$ structures still present in the image [3]. Assuming a Gaussian model for the remaining residuals is not enough to reach the expected photon-limited precision, especially between $1.2\text{-}2.8 \mu\text{m}$ [3]. In the context of exoplanet transit spectroscopy, this spectral region is crucial for the detection of molecules such as H_2O , N_2 , CO_2 , and CH_4 [4, 5]. Recently, machine learning methods have been proposed to denoise astronomical data [6], however, existing techniques only provide point estimates of denoised images, leaving out uncertainties and limiting their applicability to real data.

This prompts a search for probabilistic techniques capable of learning the underlying noise distribution. In this work, we consider Score-Based diffusion Models (SBMs) [7, 8]. This framework has been shown to be effective in modeling non-Gaussian noise in telescope images [9], since an SBM trained using noise patches can be used as a likelihood model. When combined with a prior SBM in a Bayesian inference setting, we show that our framework is able to perform denoising on spectral traces from the JWST NIRISS instrument.

2 Methods

2.1 Score-based models for linear inverse problems

Focusing on the purely additive noise component, we can cast the problem of denoising a spectral trace as a special case of a linear inverse problem with the following data generation process

$$\mathbf{y} = A\mathbf{x} + \boldsymbol{\eta} \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^d$ is a vector containing the pixel values of an observed spectral trace from the NIRISS instrument, $\mathbf{x} \in \mathbb{R}^d$ are the parameters of interest — the pixel values of an image of the noiseless spectral trace — and $\boldsymbol{\eta} \in \mathbb{R}^d$ is a random variable that represents additive noise corrupting the observation. A is the forward model applied to the noiseless spectral trace. Throughout this work, we use the identity $A \equiv \mathbb{1}_{d \times d}$ as our forward model.

Our goal is to recover the parameters of interest, \mathbf{x} , from the corrupted observation, \mathbf{y} . In this work, we model \mathbf{x} as an image — a 2-dimensional grid of pixel values. Focusing on the spectral traces only limits the scope of this study strictly to the denoising stage of the preprocessing pipeline of spectral traces [10], which is the first step towards conducting atmospheric retrievals.

Equation (1) becomes an ill-posed inverse problem for our purpose of denoising a spectral trace. This is due to the fact that the distribution from which $\boldsymbol{\eta}$ is sampled is not analytically defined, owing to the non-Gaussian nature of the $1/f$ noise. To solve this problem, we aim to recover the posterior distribution over noiseless spectral traces, $p(\mathbf{x} | \mathbf{y})$, in a framework given by Bayes' theorem

$$p(\mathbf{x} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x})p(\mathbf{x}), \quad (2)$$

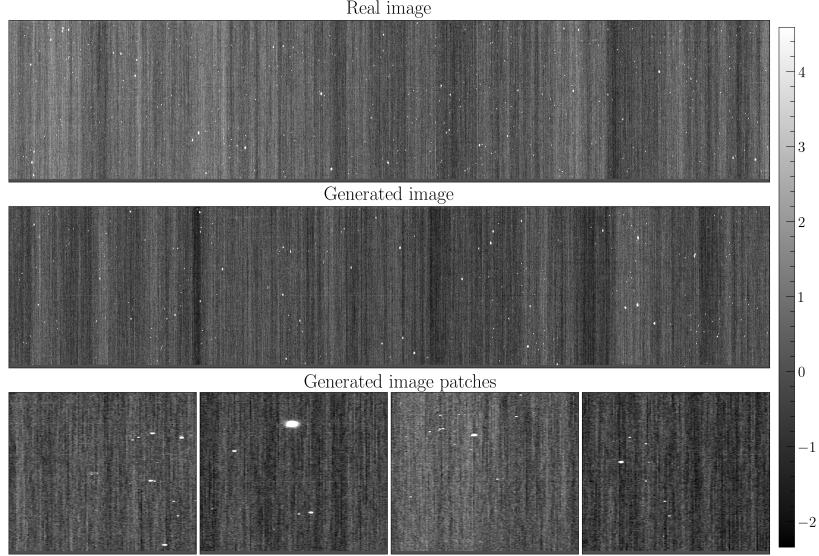


Figure 2: (Top) A real noise sample from the NIRISS darks. (Middle) A generated noise sample. (Bottom) Close-up 128x128 patches of various generated samples. The leftmost patch contains an epoxy void patch (small ring-like structure). The patch second from the left contains a bright cosmic ray. These features, along with the $1/f$ noise, have been learned by the model.

where $p(\mathbf{x})$ is the prior distribution and $p(\mathbf{y} | \mathbf{x})$ is the likelihood distribution defined by the noise model from which $\boldsymbol{\eta}$ is sampled, $q(\boldsymbol{\eta})$.

The prior and noise models are characterized using the Stochastic Differential Equation (SDE) formalism for diffusion models [7]. More specifically, we train U-net [11] neural networks $\mathbf{s}_\theta(t, \mathbf{x}) : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ to learn the score of the prior, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, and the score of the noise model, $\nabla_{\boldsymbol{\eta}_t} \log q_t(\boldsymbol{\eta}_t)$ with the denoising score-matching loss [12, 13] using the `score-models`¹ package.

To solve the inverse problem (1), we use the trained SBMs as in [14] to solve the reverse-time SDE with time-index t and with the posterior $p_t(\mathbf{x}_t | \mathbf{y})$ as its $t = 0$ boundary condition. Using Bayes' theorem, the posterior score can be written as a sum of the prior and likelihood scores

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t | \mathbf{y}) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} | \mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t). \quad (3)$$

The second term on the RHS is the score of the prior. It can be learned directly as described above using score-matching and a training dataset of simulated, noiseless spectral traces (see section 2.3).

For the likelihood score, we use the SLIC framework [9] to rewrite the likelihood score in terms of the score of the noise model via the chain rule at $t = 0$. For $t > 0$, we use the convolved likelihood approximation [14–17] to write

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} | \mathbf{x}_t) \approx -\nabla_{\boldsymbol{\eta}_t} \log q_t(\boldsymbol{\eta}_t) A. \quad (4)$$

This redefinition makes use of the linearity of the data generating process (1) w.r.t. to \mathbf{x}_t , and the fact that noise is additive, to reinterpret $\boldsymbol{\eta}_t = \mathbf{y} - \mathbf{x}_t$ as the residuals between the observation and the model. Crucially, with this redefinition we can use the score-matching procedure mentioned above to train an SBM on a dataset of noise samples described in section 2.2 to represent the likelihood.

2.2 The noise dataset

To train the noise model, we construct a training set using time-series observations of dark calibration images obtained during the commissioning phase of the NIRISS instrument, which can be found in the MAST² portal. These images were taken in the SUB256 mode as part of the NIS-006 program [18], with each observation consisting of 30 integrations comprising 30 up-the-ramp reads with a size of 256x2048 pixels. We also apply minimal preprocessing consisting of removing the superbias and constructing a bad pixel map using the `jwst` data reduction package [19].

¹github.com/AlexandreAdam/score_models

²MAST

870 spectral trace images per time-series observation are obtained by subtracting read 1 from read 2, read 3 from read 4, and so on, assuming the individual reads are approximately independently and identically distributed (iid). This ensures that training images capture the accumulation rate of cosmic rays. With eight time-series dark observations, this yields 6960 individual dark images, of which 4000 are randomly selected to train the noise model due to the high dimensionality of the images and the size of the U-nets, which constrains the number of images that can fit on one GPU. This method assumes that there are no significant temporal correlations from one dark read to the next. A representative example from the true noise training set is displayed in panel 1 of Figure 2.

2.3 Stellar trace simulations

To train the prior model, we generate simulations of noiseless spectral traces. Specifically, we use PHOENIX stellar models [20] ranging from stars with surface temperatures between $2400K$ to $6900K$, and produce simulations of traces by convolving these spectra with real PSFs from the NIRISS instrument acquired between June and August 2022. The PSFs were simulated with WebbPSF [21], using optical path difference (OPD) maps produced by the JWST wavefront sensor. To achieve a diverse training set, we also vary the horizontal shift, the rotation of the trace on the detector, and the brightness of the underlying trace. This results in 19,035 training examples for the prior model (see Figure 3).

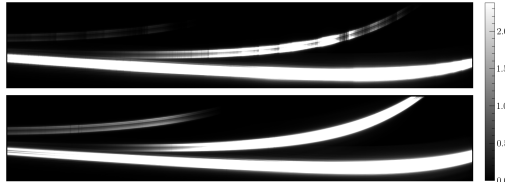


Figure 3: Examples of stellar trace simulations used to train the prior model. The first order trace is the brightest spectral trace at the bottom, ranging from $0.84\mu\text{m}$ to $2.83\mu\text{m}$, while the second order trace ranges from $\sim 0.6\mu\text{m}$ to $0.82\mu\text{m}$ [3]. A very faint third order trace is also visible at the top. Absorption lines differ depending on the stellar model used — the top and bottom traces were simulated for stars at $T = 2800K$ and $T = 6800K$ respectively.

3 Results and discussion

3.1 Noise model

We first visually assess the accuracy of the trained noise model. Figure 2 compares a SBM-generated noise sample with a sample from the real JWST darks. To make the comparison easier, we also display zoom-ins of different regions of the generated darks. Samples from the noise model can reproduce all the features of the various sources of noise in dark images, including epoxy void patches, $1/f$ noise, dead pixels, and cosmic rays. To quantitatively assess the statistical accuracy of the model, we perform the PQMass test [22], which is a density-free, sample-based method for assessing the quality of generative models. This approach enables the estimation of the probability that two sets of samples are drawn from the same distribution. This comparison can be conducted by dividing the space into non-overlapping regions and comparing the number of data samples in each region. Since it can be applied directly to high-dimensional data, it obviates the need for dimensionality reduction typically required in quality assessment metrics. In our case, we use 10 regions and compute the probability that 2000 real darks and 2000 generated samples are drawn from the same underlying data-generating process. We obtain a χ^2_{PQM} of 43, which is within 2σ of the expected value of 9, testifying to the quality of the over 500,000-dimensional samples generated by our trained noise model.

3.2 Inference

To evaluate the performance of our inference methodology for a spectral trace observed with the NIRISS instrument, we sample the underlying trace from the mock observation in Figure 1. The mock observation was constructed by overlaying a test simulation made from a PHOENIX stellar model at $T = 2400K$ with a real dark image. Figure 1 shows that our method recovers the spectrum well, with clearer features than those seen in the clean spectrum reconstructed with the state-of-the-art method. The residual shows faint artifacts along the location of the first- and second-order spectral traces, which we suspect are due to the prior model not representing the distribution of simulations sufficiently well.

To further test the performance of our noise correction technique, we extract the 1-dimensional spectrum of the star using the ground truth trace, the state-of-the-art reconstructed traces, and the 100 posterior samples. We first perform the state-of-the-art $1/f$ and cosmic ray correction on 100 mock observations constructed using the same ground truth spectral trace overlaid with 100 different real

dark realizations. Then, we fit the three orders of the trace for the ground truth, median of the 100 mock observations, and median of the 100 posterior samples, and finally perform box extractions using the `jwst` package [19]. This results in the extraction of the 1-dimensional spectrum of the star for each of the three spectral traces. Figure 4 highlights the median extracted fluxes in arbitrary units (fluxes were normalized between 0 and 1 for ease of display), as well as the standard deviation comparison between the state-of-the-art denoising method and our own. The pixel range shown reflects the wavelength regions of highest concern, from 2.4 to 2.84 μm , where the excess noise is 1.3-1.5 times higher than expected under the state-of-the-art method [3].

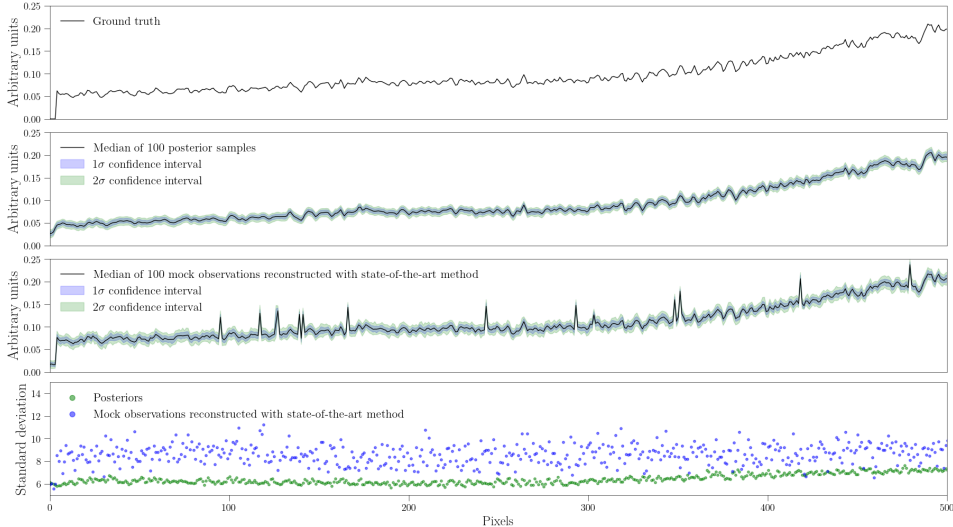


Figure 4: (Top) The extracted 1-d spectrum of the ground truth spectral trace. (Middle top) The extracted 1-d spectrum of the median of our 100 posterior samples, with the 1σ and 2σ uncertainties. (Middle bottom) The extracted 1-d spectrum of the median of 100 mock observations reconstructed with the state-of-the-art method, with the 1σ and 2σ uncertainties. (Bottom) The standard deviations of the median posterior extracted spectrum (green) and the median state-of-the-art extracted spectrum (blue). With median standard deviations of 9.0 and 6.6 for the state-of-the-art method and our method respectively, this shows a 36% increase in precision within the 1-500 pixel range, corresponding to the wavelength range of 2.4-2.8 μm .

In the last panel of Figure 1, we display the pixel-wise 85% – 15% confidence interval of the 100 posterior reconstructions to represent their uncertainties. These emphasize that the very bright regions of the trace are modeled with larger uncertainties. Given that the noise structures are absent from the reconstructions and the high performance of the noise model under the PQMass test, the bright region in the confidence interval map alludes to our prior model being misspecified and thus poorly representing the more extreme values in the pixel distribution. We further note that, due to the very large dimensionality of the images (256x2048), our architecture might have suboptimal inductive biases and benefit from hyperparameter fine-tuning tailored for the large-scale correlation structures found in our trace simulations. We leave this to future work.

Our work showcases the potential of score-based likelihoods and priors in mitigating $1/f$ noise in spectroscopic data. This proof-of-concept used a simplified model without considering up-the-ramp accumulation or Poisson noise. Future work will incorporate these elements to accurately model individual spectral trace reads. Applied across a time-series of traces, this approach could significantly enhance transit spectroscopy sensitivity, advancing exoplanet atmosphere characterization.

Acknowledgements

We would like to thank Ronan Legin for useful discussions on the use of SLIC for our application.

S.S. was supported through the Canadian Graduate Scholarships — Master’s program (CGS M) provided by the National Science and Engineering Research Council (NSERC). Compute resources were provided by the Digital Research alliance of Canada (<https://www.alliancecan.ca/en>) and Calcul Québec (<https://www.calculquebec.ca/>).

References

- [1] Armin Rest. Readnoise and 1/f-noise characterization of the NIRCcam detectors. Technical Report JWST-STScI-004118, March 2015.
- [2] René Doyon, Chris J. Willott, John B. Hutchings, Anand Sivaramakrishnan, Loïc Albert, David Lafrenière, Neil Rowlands, M. Begoña Vila, André R. Martel, Stephanie LaMassa, David Aldridge, Étienne Artigau, Peter Cameron, Pierre Chayer, Neil J. Cook, Rachel A. Cooper, Antoine Darveau-Bernier, Jean Dupuis, Colin Earnshaw, Néstor Espinoza, Joseph C. Filippazzo, Alexander W. Fullerton, Daniel Gaudreau, Roman Gawlik, Paul Goudfrooij, Craig Haley, Jens Kammerer, David Kendall, Scott D. Lambros, Luminita Ilinca Ignat, Michael Maszkiewicz, Ashley McColgan, Takahiro Morishita, Nathalie N. Q. Ouellette, Camilla Pacifici, Natasha Philippi, Michael Radica, Swara Ravindranath, Jason Rowe, Arpita Roy, Niladri Roy, Karl Saad, Sangmo Tony Sohn, Geert Jan Talens, Driss Touahri, Deepashri Thatte, Joanna M. Taylor, Thomas Vandal, Kevin Volk, Michel Wander, Gerald Warner, Sheng-Hai Zheng, Julia Zhou, Roberto Abraham, Mathilde Beaulieu, Björn Benneke, Laura Ferrarese, Ray Jayawardhana, Doug Johnstone, Lisa Kaltenegger, Michael R. Meyer, Judy L. Pipher, Julien Rameau, Marcia Rieke, Salma Salhi, and Marcin Sawicki. The Near Infrared Imager and Slitless Spectrograph for the James Webb Space Telescope. I. Instrument Overview and In-flight Performance. *Publ. Astron. Soc. Pac.*, 135(1051):098001, September 2023.
- [3] Loïc Albert, David Lafrenière, René Doyon, Étienne Artigau, Kevin Volk, Paul Goudfrooij, André R. Martel, Michael Radica, Jason Rowe, Néstor Espinoza, Arpita Roy, Joseph C. Filippazzo, Antoine Darveau-Bernier, Geert Jan Talens, Anand Sivaramakrishnan, Chris J. Willott, Alexander W. Fullerton, Stephanie LaMassa, John B. Hutchings, Neil Rowlands, M. Begoña Vila, Julia Zhou, David Aldridge, Michael Maszkiewicz, Mathilde Beaulieu, Neil J. Cook, Caroline Piaulet, Pierre-Alexis Roy, Pierrot Lamontagne, Kim Morel, William Frost, Salma Salhi, Louis-Philippe Coulombe, Björn Benneke, Ryan J. MacDonald, Doug Johnstone, Jake D. Turner, Marylou Fournier-Tondreau, Romain Allart, and Lisa Kaltenegger. The Near Infrared Imager and Slitless Spectrograph for the James Webb Space Telescope. III. Single Object Slitless Spectroscopy. *Publ. Astron. Soc. Pac.*, 135(1049):075001, July 2023.
- [4] Nikku Madhusudhan, Subhjit Sarkar, Savvas Constantinou, Måns Holmberg, Anjali Piette, and Julianne I. Moses. Carbon-bearing Molecules in a Possible Hycean Atmosphere. *ArXiv.org*, art. arXiv:2309.05566, September 2023.
- [5] Charles Cadieux, René Doyon, Ryan J. MacDonald, Martin Turbet, Étienne Artigau, Olivia Lim, Michael Radica, Thomas J. Fauchez, Salma Salhi, Lisa Dang, Loïc Albert, Louis-Philippe Coulombe, Nicolas B. Cowan, David Lafrenière, Alexandrine L’Heureux, Caroline Piaulet-Ghorayeb, Björn Benneke, Ryan Cloutier, Benjamin Charnay, Neil J. Cook, Marylou Fournier-Tondreau, Mykhaylo Plotnykov, and Diana Valencia. Transmission Spectroscopy of the Habitable Zone Exoplanet LHS 1140 b with JWST/NIRISS. *Astrophys. J. Lett.*, 970(1):L2, July 2024.
- [6] Guillaume Payeur, Étienne Artigau, Laurence Perreault Levasseur, and René Doyon. Correlated Read Noise Reduction in Infrared Arrays Using Deep Learning. *Astron. J.*, 163(6):292, June 2022.
- [7] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv e-prints*, art. arXiv:2011.13456, November 2020.
- [8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.
- [9] Ronan Legin, Alexandre Adam, Yashar Hezaveh, and Laurence Perreault-Levasseur. Beyond Gaussian Noise: A Generalized Approach to Likelihood Analysis with Non-Gaussian Noise. *Astrophys. J. Lett.*, 949(2):L41, June 2023.
- [10] Howard Bushouse, Jonathan Eisenhamer, Nadia Dencheva, James Davies, Perry Greenfield, Jane Morrison, Phil Hodge, Bernie Simon, David Grumm, Michael Droettboom, Edward

- Slavich, Megan Sosey, Tyler Pauly, Todd Miller, Robert Jedrzejewski, Warren Hack, David Davis, Steven Crawford, David Law, Karl Gordon, Michael Regan, Mihai Cara, Ken MacDonald, Larry Bradley, Clare Shanahan, William Jamieson, Mairan Teodoro, and Thomas Williams. Jwst calibration pipeline, October 2022.
- [11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [12] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.
- [13] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Comput.*, 23(7):1661–1674, jul 2011. ISSN 0899-7667.
- [14] Alexandre Adam, Connor Stone, Connor Bottrell, Ronan Legin, Yashar Hezaveh, and Laurence Perreault-Levasseur. Echoes in the Noise: Posterior Samples of Faint Galaxy Surface Brightness Profiles with Score-Based Likelihoods and Priors. *arXiv e-prints*, art. arXiv:2311.18002, November 2023.
- [15] B. Remy, F. Lanusse, N. Jeffrey, J. Liu, J. L. Starck, K. Osato, and T. Schrabback. Probabilistic mass-mapping with neural score estimation. *A&A*, 672:A51, April 2023.
- [16] Alexandre Adam, Adam Coogan, Nikolay Malkin, Ronan Legin, Laurence Perreault-Levasseur, Yashar Hezaveh, and Yoshua Bengio. Posterior samples of source galaxies in strong gravitational lenses with score-based priors. In *Machine Learning and the Physical Sciences Workshop*, page E1, January 2022.
- [17] Noe Dia, M. J. Yantovski-Barth, Alexandre Adam, Micah Bowles, Pablo Lemos, Anna M. M. Scaife, Yashar Hezaveh, and Laurence Perreault-Levasseur. Bayesian Imaging for Radio Interferometry with Score-Based Priors. *arXiv e-prints*, art. arXiv:2311.18012, November 2023.
- [18] André M. Martel and Niriss Team. A Summary of the Commissioning of JWST NIRISS, March 2023.
- [19] H. Bushouse. The JWST Science Calibration Pipeline. In R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, and H. Verkouter, editors, *Astronomical Data Analysis Software and Systems XXIX*, volume 527 of *Astronomical Society of the Pacific Conference Series*, page 583, January 2020.
- [20] T. O. Husser, S. Wende-von Berg, S. Dreizler, D. Homeier, A. Reiners, T. Barman, and P. H. Hauschildt. A new extensive library of PHOENIX stellar atmospheres and synthetic spectra. *Astron. Astrophys.*, 553:A6, May 2013.
- [21] Marshall D. Perrin, Joseph Long, Anand Sivaramakrishnan, Charles-Phillipe Lajoie, Erin Elliot, Laurent Pueyo, and Loic Albert. WebbPSF: James Webb Space Telescope PSF Simulation Tool. *Astrophysics Source Code Library*, record ascl:1504.007, April 2015.
- [22] Pablo Lemos, Sammy N. Sharief, Nikolay Malkin, Laurence Perreault-Levasseur, and Yashar D. Hezaveh. Pqmass: Probabilistic assessment of the quality of generative models using probability mass estimation. *ArXiv*, abs/2402.04355, 2024.