Research in Astronomy and Astrophysics, 25:035009 (14pp), 2025 March

https://doi.org/10.1088/1674-4527/adb13a
CSTR: 32081.14.RAA.adb13a

Tests of Solar X-Ray Image Reconstruction: A New Index for Assessing Image Quality

Zhen-Tong Li[1], Wen-Hui Yu[1,2], Yang Su[1,2], Wei Chen[1,2], and Wei-Qun Gan[1,3]
[1] Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023, China; ztli@pmo.ac.cn, yang.su@pmo.ac.cn
[2] School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, China
[3] University of Chinese Academy of Sciences (UCASNJ), Nanjing 211135, China
Received 2024 November 19; revised 2025 January 20; accepted 2025 January 22; published 2025 March 10

Abstract

Indirect X-ray modulation imaging has been adopted in a number of solar missions and provided reconstructed X-ray images of solar flares that are of great scientific importance. However, the assessment of the image quality of the reconstruction is still difficult, which is particularly useful for scheme design of X-ray imaging systems, testing and improvement of imaging algorithms, and scientific research of X-ray sources. Currently, there is no specified method to quantitatively evaluate the quality of X-ray image reconstruction and the point-spread function (PSF) of an X-ray imager. In this paper, we propose percentage proximity degree (PPD) by considering the imaging characteristics of X-ray image reconstruction and in particular, sidelobes and their effects on imaging quality. After testing a variety of imaging quality assessments in six aspects, we utilized the technique for order preference by similarity to ideal solution to the indices that meet the requirements. Then we develop the final quality index for X-ray image reconstruction, QuIX, which consists of the selected indices and the new PPD. QuIX performs well in a series of tests, including assessment of instrument PSF and simulation tests under different grid configurations, as well as imaging tests with RHESSI data. It is also a useful tool for testing of imaging algorithms, and determination of imaging parameters for both RHESSI and ASO-S/Hard X-ray Imager, such as field of view, beam width factor, and detector selection.

Key words: Sun: flares – Sun: X-rays, gamma-rays – techniques: image processing

1. Introduction

Solar X-ray flares are one of the most powerful solar activities. The existence of X-ray flares implies strong plasma heating and particle acceleration processes, an essential subject for understanding energy release and transport mechanisms (e.g., Hudson & Ryan 1995; Holman et al. 2011). However, it is difficult to directly image X-ray sources in hard X-rays using an optical focusing system. Indirect image techniques such as Fourier transform imaging systems are often used to observe and reconstruct X-ray solar images (Hurford 2010).

A typical system of such a technique consists of a series of subcollimators: pairs of co-aligned grids with different position angles, pitches, and phases, which are utilized in the many previous X-ray missions, e.g., the Hard X-ray Telescope (HXT. Kosugi (Kosugi et al. 1991)) onboard Yohkoh (Ogawara et al. 1991), the RHESSI (Lin et al. 2002) and the Spectrometer Telescope for Imaging X-rays onboard Solar Orbiter (STIX; Krucker et al. 2020). The Hard X-ray Imager (HXI; Su et al. 2019, 2024; Zhang et al. 2019) aboard the Advanced Space-based Solar Observatory (ASO-S; Gan et al. 2019, 2023) is so far the most recent high-energy X-ray imaging instrument using the spatial-modulation Fourier transform method (Su et al. 2019).

It is known that the quality of a reconstructed X-ray image is usually not as good as that of direct imaging in terms of details and dynamic range due to the limited number of subcollimators or visibilities (i.e., Fourier components of the X-rays sources sampled in specific $u-v$ points linked to subcollimators characteristics). The assessment of imaging quality is therefore important for the scheme design of an imaging system, the testing and improvement of imaging algorithms, as well as scientific research.

However, quantitative evaluation of image quality of reconstructed X-ray images from modulation-based imager has been a difficult work. Several previous studies mainly concentrate on measuring the difference between the reconstructed image and the observed signal to evaluate the X-ray imaging goodness, such as the C-statistic and χ^2 test (e.g., Dennis & Tolbert 2019; Massa et al. 2022). Besides, there is no specified means to evaluate the point-spread function (PSF) of an X-ray imager, where the amount and intensity of sidelobes should be considered as a crucial factor. On the other hand, methods for evaluating image quality in various aspects, referred to as the image quality assessment methods, have been developed and widely investigated in the computing and engineering field, especially over the last two decades (e.g.,

Table 1
Existent Image Quality Assessment Indices in Image Engineering Fields

	General	Specific
FR	Root mean squared error (RMSE)	...
	Structural similarity index measure (SSIM)	...
	Cross entropy (CE)	Gradient-based fusion performance ($Q^{AB/F}$)
	Mutual information (MI)	Chen–Blum metric (Q_{CB})
	Peak signal-to-noise ratio (PSNR)	Chen–Varshney metric (Q_{CV})
	Information content weighted SSIM (IW-SSIM)	Visual information fidelity (VIF)
	Multi-Scale SSIM (MS-SSIM)	Information fidelity criterion (IFC)
	Feature similarity (FSIM)	...
NR	Entropy (EN)	...
	Average gradient (AG)	...
	Edge intensity (EI)	...
	Standard deviation (SD)	...
	Spatial frequency (SF)	...

Note. FR: full-reference. NR: no-reference.

Wang & Li 2011; Islam et al. 2013). As introduced in Sara et al. (2019), the objective image quality merits are classified mainly into two groups: full-reference measurements and no-reference measurements. No-reference metrics only calculate different parameters of a test image itself, but full-reference metrics can describe the goodness of a reconstructed or distorted image by directly comparing it with the original image that must be accessible. When evaluating an imaging system, full-reference metrics are more suitable for measuring the similarity or difference between simulated sources and reconstructed sources. In the image fusion field, Liu et al. (2012) provided 12 evaluation metrics to compare fused image and original image, and those metrics are also utilized in evaluating the image fusion benchmark in Zhang et al. (2020).

The evaluation metrics mentioned above have rarely been utilized in quality assessment of reconstructed X-ray image. One of the reasons is that it is difficult to describe the various special features related to X-ray image reconstruction. Chen et al. (2020) used intensity evolution curves of side lobe in reconstructed X-ray images under different grid configuration models (different number of and combinations of grid pitches and position angles) to evaluate their performance. It is logical that the grid configuration is better when the sidelobes in the PSF are less and weaker. Nevertheless, no quantitative method has been established to effectively compare the curves in that work.

In this study, we are aiming to establish a quantitative, effective index to describe the quality of reconstructed X-ray images, including instrument PSF. We first survey various existing image quality assessment methods that describe similarity of images, design the new index taking into account the feature characteristics of reconstructed X-ray images, and test the indices in Section 2. In Section 3, we establish the quality index for X-ray image reconstruction (henceforth QuIX). The test results are presented in Section 4. Several applications of QuIX in X-ray imaging are presented in Section 5. A brief summary and a discussion are given in Section 6.

2. Image Quality Assessment Indices

For the purpose of quantitatively and effectively describing the quality of reconstructed X-ray images, the existing image quality assessment methods need to be considered. Many of them constitute the benchmarks in the image fusion and the image reconstruction fields (Sara et al. 2019; Zhang et al. 2020). But as introduced, specified methods also need to be developed to effectively describe the features often seen in reconstructed X-ray images, natural or artificial, such as sidelobes. In the following sub-sections, we will test the generally adopted indices and develop the new index for X-ray image reconstruction.

2.1. Existent Image Quality Assessment Indices

The image quality assessment methods can be grouped into two categories as full-reference and no-reference (Sara et al. 2019), based on the availability of reference images. The full-reference methods require at least one reference image to compute the similarity between, and the no-reference approaches only calculate the parameters from the image itself. As already summarized in Zhang et al. (2020), there are 12 evaluation metrics to evaluate the quality of the fused images, including both full-reference and no-reference metrics. Most of those metrics can be generally applied to other types of images, but a few of them are specially designed for fused images. Table 1 shows a quick look and simple division of existent image quality assessment indices summarized in Zhang et al. (2020), gathering with some other methods developed later such as information content weighted SSIM (IW-SSIM; Wang & Li 2011), multi-scale SSIM (MS-SSIM; Wang et al. 2003),

feature similarity (FSIM; Zhang et al. 2011), and the metrics constructed for images of natural scene statistics that correlate well with human visual system such as visual information fidelity (Sheikh & Bovik 2006) and information fidelity criterion (Sheikh et al. 2005), etc.

To evaluate the reconstructed X-ray image quality, the comparison between the reconstructed image and the ground truth is needed in this work, rather than evaluating the reconstructed image itself. Thus the no-reference approach does not apply to this work. The specifically designed indices in Table 1 are also not proper to be applied, since the emphasis on the quality of the reconstructed images is different from that of natural scene statistics images, and those indices designed for multiple images fusion are inapplicable to the double-object cases. Therefore, in summary, there are eight indices that are applicable in this study, i.e., RMSE, SSIM, CE, MI, PSNR, IW-SSIM, MS-SSIM and FSIM. It should be noted that these eight indices still need further tests to exclude improper or unstable evaluation for the reconstruction quality of X-ray images, which is given in Section 2.3.

2.2. Development of Percentage Proximity Degree

We first consider sidelobes, one of the most crucial characteristics in X-ray image reconstruction that affect the overall image quality, especially the map generated through back projection (BP) approach. In principle, sidelobes should be minimized in both amount and strength, which is helpful to avoid generating or reduce artificial structures. The intensity of sidelobes is in general smaller than the peak intensity of the source itself, and the number of sidelobes usually increases with decreasing intensity threshold of the BP map. Thus we consider the number of pixels above a threshold of level p (%) of the maximum intensity of the map $N_{\mathrm{BP,cts}>p}$, and take the ratio of $N_{\mathrm{BP,cts}>p}$ to the total number of pixels N_0 of the map as

$$A_{\mathrm{BP}}(p) = \frac{N_{\mathrm{BP,cts}>p}}{N_0}. \quad (1)$$

For the original X-ray source, the ratio is written as $A_s(p)$. The difference between $A_s(p)$ and $A_{\mathrm{BP}}(p)$ should be smaller at level p (%) for a BP map with better image quality. Thus it is reasonable to establish the proximity degree between the two ratios as functions of p to evaluate the contribution of sidelobes, which expresses the image quality to a certain extent. Here we define the percentage proximity degree (PPD) as

$$\begin{aligned}\mathrm{PPD} &= \int_a^b [A_{\mathrm{BP}}(p) - A_s(p)]dp \\ &= \int_a^b \frac{N_{\mathrm{BP,cts}>p} - N_{\mathrm{s,cts}>p}}{N_0}dp, \end{aligned} \quad (2)$$

where a and b are the integral limits, and the default value is 1 and 100, respectively. Equation (2) describes the integral of the difference between $A_s(p)$ and $A_{\mathrm{BP}}(p)$ (i.e., the blue area between the two black curves in Figure 1(c)). Apparently, the value of PPD increases when the BP map has more heavy sidelobes.

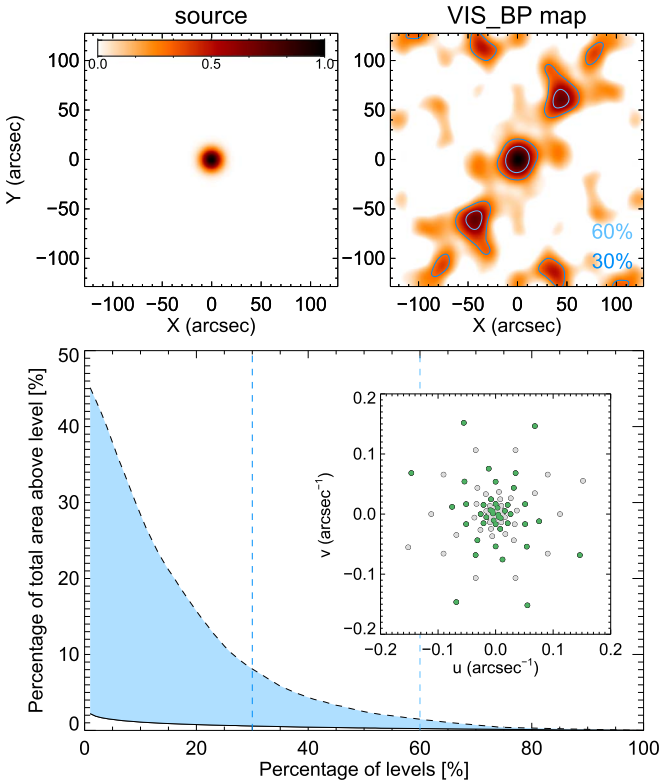

Figure 1. An example of the Percentage Proximity Degree (PPD) for a simulated case with a single source. Panel (a): Simulated Gaussian source with $\sigma = 5''$. Panel (b): the reconstructed map of the Gaussian source using VIS_BP algorithm. The dark blue and light blue contours correspond to 30% and 60% levels, respectively. Panels (a) and (b) share the same color bar. Panel (c): The dashed black curve is $A_{\mathrm{BP}}(p)$, the solid black curve is $A_s(p)$. The light blue area in between is the PPD. The two vertical dashed blue lines correspond to the contours in panel (b). The green points in the inset plot indicate the $u-v$ pairs that are used to generate the VIS_BP map.

2.3. Testing Indices

It is necessary to select proper indices after introducing the eight applicable image quality assessment indices in Section 2.1 and the new one, PPD in Section 2.2. For quantification and convenience, we assume that the reconstructed X-ray image should have specific features in five different aspects, which are listed below. Therefore, it is presumable/conceivable that the image quality assessment indices should correctly respond to changes of those features.

(1) *Consistency*. The imaging quality should be improved when the number of $u-v$ pairs (grid configuration

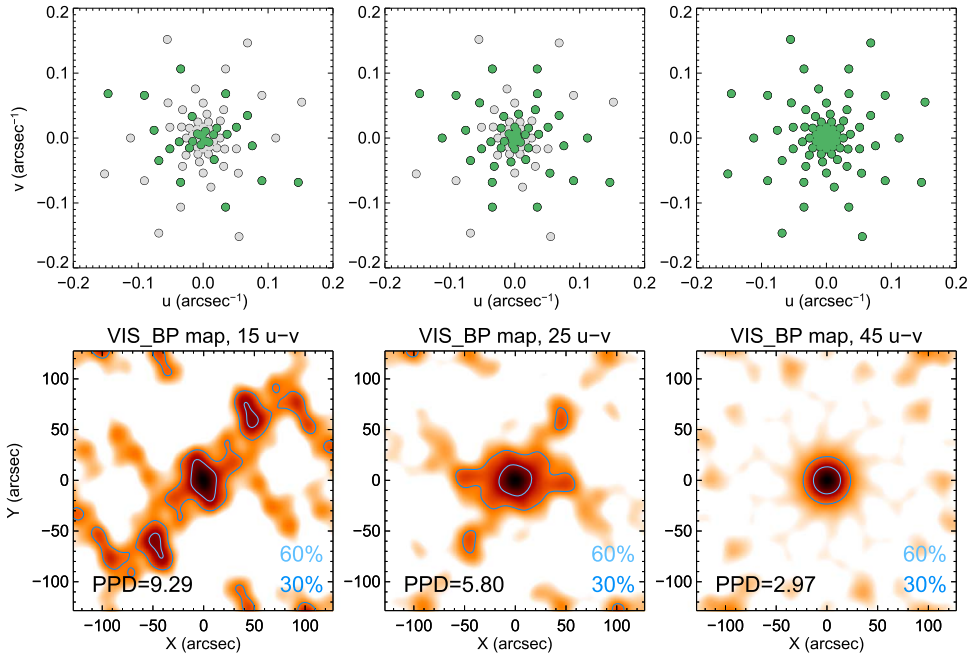

Figure 2. The performance of PPD with varying $u-v$ number. Top panels: $u-v$ maps. The green dots represent the $u-v$ configurations used to make the VIS_BP maps. All the hollow points in each map represent the $u-v$ of HXI grid configuration. Bottom panels: corresponding VIS_BP maps. Each map is reconstructed from a simulated image of a single Gaussian source with $\sigma = 5''$ and the corresponding $u-v$ numbers: 15, 25 and 45 pairs, respectively. The percentage proximity degree PPD is marked in each map.

parameters) increases (of course, how $u-v$ distributes is also important).

(2) *Shape*. The imaging quality should get worse when the shape of the reconstructed image is more distorted.

(3) *Size*. The imaging quality should deteriorate as the size difference between the reconstructed source and the original source gets larger.

(4) *Position*. The imaging quality should deteriorate as the position of the reconstructed source is further away from the original location.

(5) *Number of sidelobes*. The imaging quality should deteriorate as more sidelobes appear in the reconstructed image.

(6) *Orientation*. The imaging quality should have no obvious change when the reconstructed image is only rotated due to the rotation of the imaging system.

Ideally, the image quality should be better if the X-ray imager configures more types of grid pitches and position angles (i.e., more $u-v$ points, more visibilities). We get a series of images that are reconstructed from a simulated Gaussian source with $\sigma = 5''$ using increasing number of $u-v$ points (from 11 to 45 pairs), to test the responses of the nine indices. As a preview of the sequence of images of increasing $u-v$ number, Figure 2 shows three samples (selected from the 11 to 45 pairs series) with $u-v$ points of 15, 25 and 45 pairs, corresponding to the three BP maps below each $u-v$ map. It is noticeable that the sidelobes are fainter and smaller as the number of $u-v$ is increasing, and the calculated PPD (marked in the bottom panels of Figure 2) also decreases. Note that the $u-v$ configurations are chosen from the complete HXI's grid configuration in Zhang et al. (2019) (all the hollow $u-v$ points in Figure 2).

After forward processing on the nine indices, all indices tend to be larger when image quality is better, and thus they should increase with the increasing number of $u-v$ pairs in this case. The results are gathered in Figure 3, where the consistency column shows the trends of index scores. The other five columns present the results of the other five test aspects, and the index scores change with the designed samples series whose shape is distorting, whose size is expanding, whose position is moving from the left to the right of the source, whose sidelobes are increasing and whose structure is rotating. According to the tendency curves of each index in each test, we can judge whether it performs as we expected. The cross marks and check marks determined from the tendency curves in Figure 3 label the indices that are beneficial to compose the final index, and that would influence the overall evaluation, respectively. Thus, there are four candidates left: PSNR, SSIM, RMSE, PPD, which performed well without any abnormal behaviors. Note that there are light cross marks in CE, IW-SSIM and PPD, indicating that they showed slight response or no response to the changes in the sources.

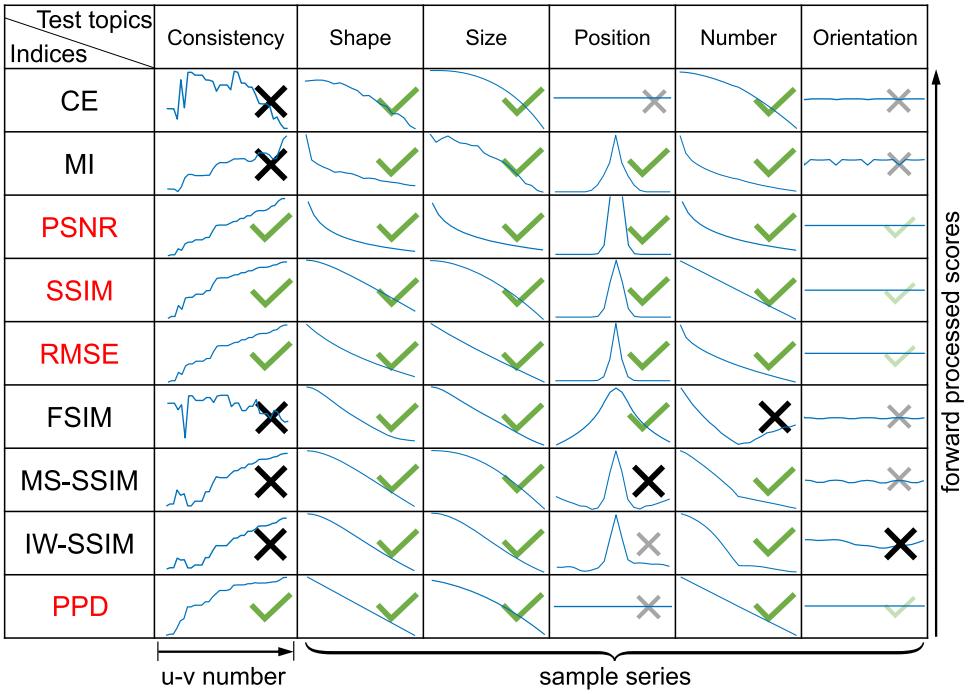

Figure 3. Synthesis of indices screening test. Each curve in each aspect of each index indicates the tendency of the index score changing with the sample series. The sample series for the consistency test consists of reconstructed BP images with increasing $u - v$ numbers. The sample series in the other aspects are reconstructed from cases with more distorted shape, increasing size, changing positions of the source from the left to the right of the original location, increasing sidelobes, and rotating orientation, respectively. The black cross marks indicate that the response of the index in this test is not satisfying our selection criteria, and the smaller gray ones indicate they had slight response or no response to the changes in the sources. The check marks in green indicate the response that meets the selection criteria, but the smaller light ones indicate that the responses are not crucial for the selection. All the index scores are forward processed.

3. The Quality Index for X-Ray Image Reconstruction

The four indices selected in Section 2.3 have different response magnitudes for different aspects. For SSIM, its larger value corresponds to better image quality, and it ranges from 0 to 1. But for RMSE, a smaller value corresponds to better image quality. The percentage proximity degree PPD ranges from $a-1$ to $b-1$. PSNR can theoretically reach infinity. Therefore, a composite indicator needs to be constructed to ensure the accuracy and validity of its response, thereby allowing for a reasonable measurement of imaging quality. Note that PSNR is defined by RMSE as $\mathrm{PSNR} = 20\log_{10}\frac{\mathrm{MAX_I}}{\mathrm{RMSE}}$ (Jagalingam & Hegde 2015), where $\mathrm{MAX_I}$ is the maximum intensity of the map. PSNR and RMSE are strongly correlated according to the equation and their tendency curves in Figure 3. Hence, we exclude PSNR from the index candidates.

Different grid configurations of an X-ray imager could result in different reconstructed maps from the same X-ray source. The best score of each evaluating index may be distributed in different maps. For instance, the image that has the least sidelobes generates the best percentage proximity degree PPD, but the image that best fits the shape of the source generates the best SSIM. The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS; Hwang & Yoon 1981) is therefore necessary to be introduced to properly order the quality of all the reconstructed images.

The overall method of TOPSIS is used to order the samples according to the distance between the sample and the optimal solution, which consists of the best score of each index in all samples. The sample which has the shortest distance from the optimal solution would be listed at the top, even though it may not contain all the optimum scores. The TOPSIS method mainly includes the following steps:

(i) Data processing, generally including forward processing and standardization;
(ii) Computing the extremum solutions, including the optimal and the worst solution;
(iii) Computing the Euclidean distances of samples from the extremum solution (the distances from the optimal and the worst solutions are d_i^+ and d_i^-, respectively);
(iv) Sorting the samples by the final score $S_i = \frac{d_i^-}{d_i^+ + d_i^-}$, calculated from distances.

Given the X-ray source and a series of reconstructed image samples, the image quality can be relatively assessed by computing their scores by the TOPSIS method. The optimal

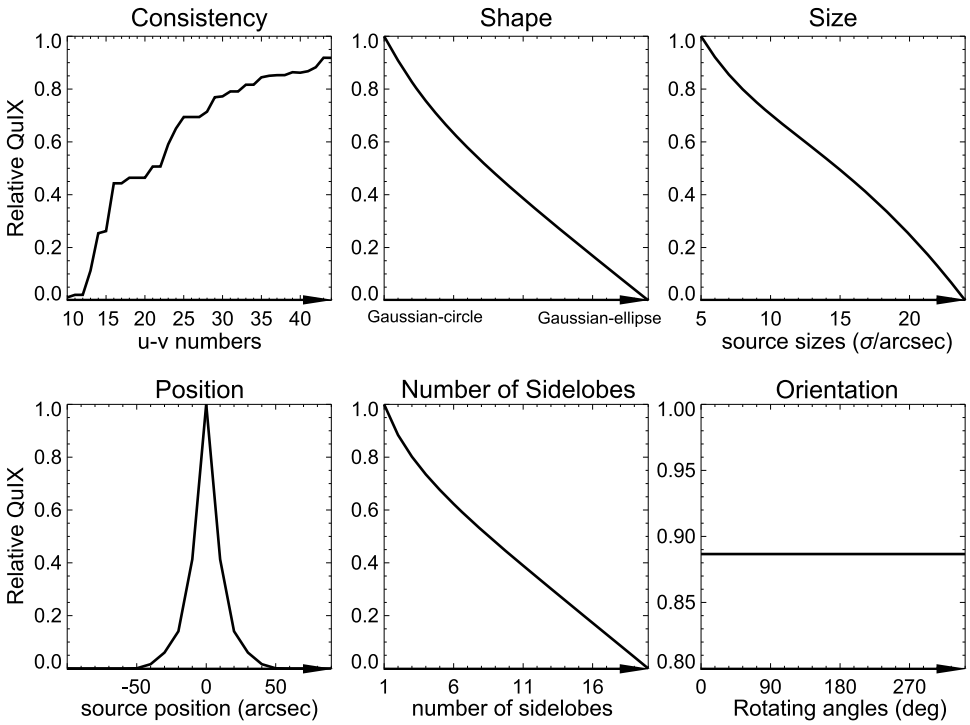

Figure 4. Test of the QuIX index for the ideal cases in the six aspects mentioned in Section 2.3. The values in *Y*-axis are normalized scores.

and the worst solutions are both from this sample series itself. When we try to assess the quality of a single reconstructed image (such as evaluation of instrument PSF), or when we want to compare the image quality among different series, the worst solution can be set as the theoretically lowest value of the three indices (i.e., $\mathrm{RMSE}^{-}=0$, $\mathrm{SSIM}^{-}=0$, $\mathrm{PPD}^{-}=0$). Similarly, the optimal solution is set to the theoretically highest value as $\mathrm{RMSE}^{+}=1$, $\mathrm{SSIM}^{+}=1$, $\mathrm{PPD}^{+}=1$. The scores calculated by this way are regarded as the absolute scores, which can be generally used to describe the image qualities in different occasions. Consequently, the final quality index for X-ray image reconstruction (QuIX) using TOPSIS method has been established, containing three indices of RMSE, SSIM and PPD.

4. Test Results

The performance of the QuIX index should be tested in the six aspects accordingly in Section 2.3, by a series of samples with corresponding features. All the calculated relative QuIX scores respond correctly for each aspect, shown in Figure 4. This basic ideal test reflects that the QuIX index could function well in representing changes of a map in these six aspects.

It is also efficient to test the performance of QuIX by evaluating the PSF or the imaging of loop sources under different grid configurations, which can help in designing the imaging system of an X-ray imager. We here rearrange the HXI grid configuration by reducing position angles (Test 1), and reducing pitch widths (Test 2,) and reconstruct images from the three different configurations using the VIS_BP algorithm. Note that the total number of $u-v$ pairs of all cases is the same as HXI. We use a point source and a simple loop source to reconstruct images through each grid configuration. The results are shown in Figure 5, with the calculated QuIX scores marked in each panel. As expected, both of the two test cases generated worse image quality than that of the HXI configuration due to the substantial side lobe artifacts in certain angles or radii (as shown in Massone et al. 2009). The QuIX scores of the maps generated by the HXI configuration are all higher (better) than those in the two test cases, which means that QuIX can correctly reflect the quality of reconstructed images. It is worth noting that, for the point source, the QuIX score of HXI is only slightly higher than Test 1 but has a larger difference with Test 2. The reason is that the pitch width of Test 1 remains the same as HXI to obtain the comparable resolution for a centrosymmetric source, although the number of position angles of Test 1 are less than that of HXI. This explanation is confirmed by the discrepancy of QuIX scores for the imaging of the loop source, which requires more position angles and pitch widths to reconstruct a resolvable image.

5. Applications

The test results of QuIX suggest that it can be used to evaluate different grid configurations of X-ray imagers, and

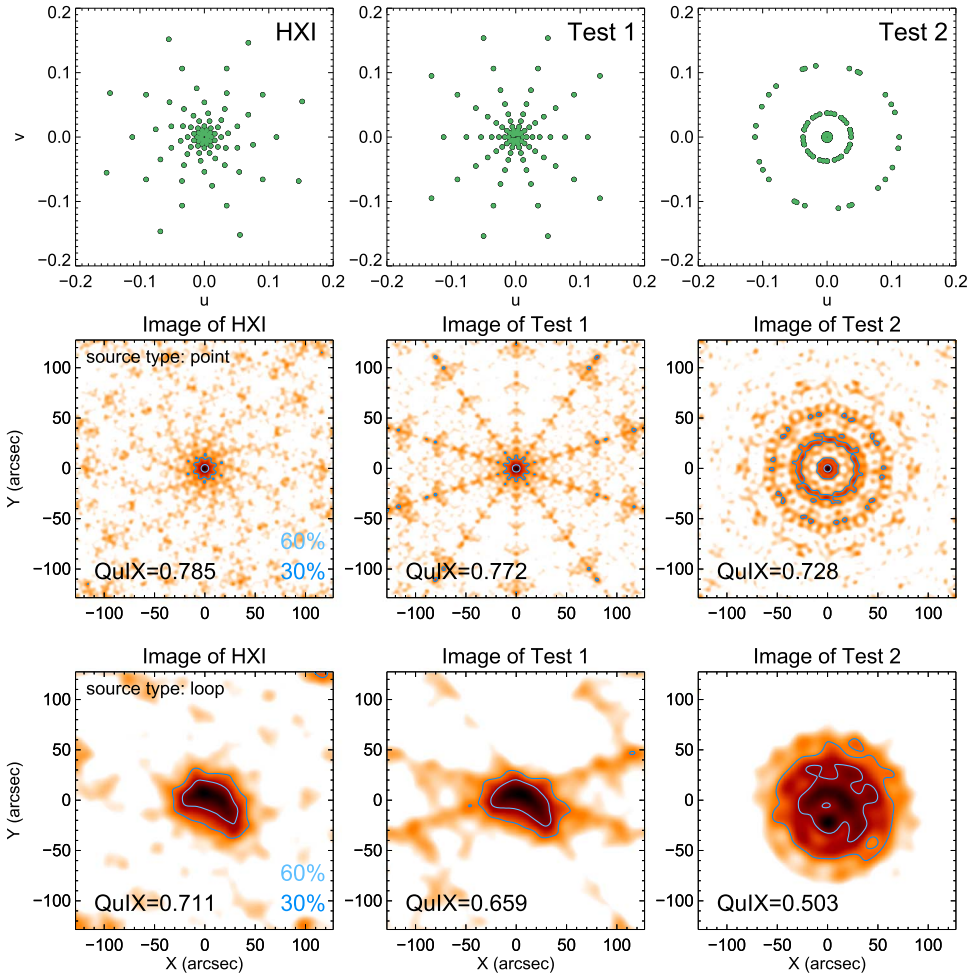

Figure 5. Test results of QuIX for different $u-v$ configurations. Top panels: $u-v$ maps of the HXI configuration and two other configurations, respectively. Middle panels: PSF maps reconstructed from the corresponding $u-v$ configuration. The QuIX score is shown in each panel. Bottom panels: reconstructed maps by the VIS_BP algorithm for the simulated case with a loop source and the corresponding $u-v$ configuration.

presumably any modulated imaging system such as radio telescopes utilizing aperture synthesis (e.g., Kellermann & Moran 2001). In this section, we present several practical applications of QuIX related to X-ray imaging processes.

5.1. Evaluation of Imaging Algorithms

Imaging algorithms play an important role in studies of solar X-ray bursts. During the RHESSI mission, over 10 imaging algorithms have been developed (see the category in the RHESSI website[4]) for different scientific applications. Therefore, it is important to evaluate the imaging results of different imaging algorithms for a better understanding of the advantages and limitations of each algorithm. Here, we do not attempt to run a full test on algorithms, but rather show the application of the QuIX index in this kind of test.

We use double Gaussian sources with $\sigma = 2''$ and same flux, which are separated by a distance of $20''$, to simulate the double footpoint sources that are frequently seen in many flare observation and simulation studies (e.g., Fletcher & Hudson 2002; Saint-Hilaire et al. 2008; Kong et al. 2022). The RHESSI simulating image routine[5] (with the real aspect data) is used to generate the reconstructed images from the algorithms CLEAN (Högbom 1974; Schwarz 1978), EM (Benvenuto et al. 2013), PIXON (Puetter 1995; Metcalf et al. 1996), Forward Fit (Aschwanden et al. 2002, 2004), MEM_NJIT (Schmahl et al. 2007), MEM_GE (Massa et al. 2020), UV_smooth (Massone et al. 2009), VIS_FWDFIT (Hurford et al. 2005), VIS_WV

[4] https://hesperia.gsfc.nasa.gov/rhessi3/software/imaging-software/image-algorithm-summary/index.html

[5] https://hesperia.gsfc.nasa.gov/rhessi3/software/simulation-software/index.html

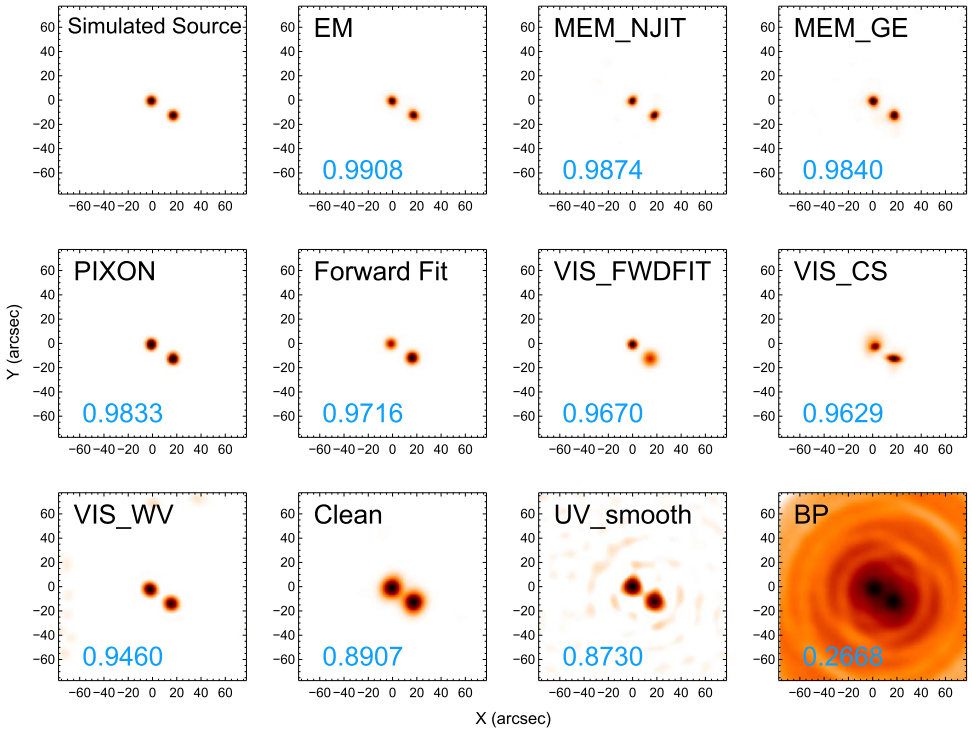

Figure 6. Testing of different imaging algorithms for the case of double Gaussian sources with the same brightness. The first panel: simulated double Gaussian sources. The rest panels: the imaging results obtained from the 11 imaging algorithms. The CLEAN beamwidth factor (CBWF) is set to 1.7, which is mentioned in Kontar et al. (2010). The QuIX scores are labeled in each panel in blue.

(Duval-Poo et al. 2018), and VIS_CS (Felix et al. 2017), respectively.

The simulated source and the images generated by the 11 imaging algorithms are shown in Figure 6 in the descending order of the QuIX scores. The EM map has the largest QuIX value and is also the closest to the model. The differences among the scores of EM, MEM_NJIT, MEM_GE, and PIXON are not significant, and they produced similar imaging results. As expected, the BP algorithm only produces basic structure of the source with the worst imaging quality. Although UV_smooth exhibits the morphology of two Gaussian sources, it also produced more sidelobes than other methods excluding BP. Both Forward Fit and VIS_FWDFIT generate a pair of sources that are different in size and intensity, resulting in lower QuIX values. The VIS_CS image has some distorted structures, and the VIS_WV sources are relatively large with slight sidelobes. These QuIX scores provide quantitative results that are more or less consistent with visual judgment.

For comparison, we also calculated the SSIM of the imaging results for each algorithm (Figure 7). The trend of SSIM is roughly consistent with QuIX, except for the order of PIXON, MEM_GE and MEM_NJIT. By examining the results of these three imaging algorithms, we found that the PIXON images show a wider structure that deviates slightly from the Gaussian source width, and that the QuIX index can effectively quantify the imaging quality of these algorithms in this test. It should be pointed out that the PPD index in QuIX mainly focuses on the sidelobes by design and therefore QuIX may be more sensitive to sidelobes than other indices. A more detailed test on these algorithms can be found in Yu et al. (2025).

To exemplify the application of the QuIX index in practical scenarios, we use it to evaluate the imaging results from actual X-ray observations. However, since the true X-ray image is usually unavailable in actual observation, it is difficult to evaluate the reconstructed images. Considering that the X-ray emission in low-energy bands (∼3–12 keV) is mainly from thermal bremsstrahlung emission, we can use the soft X-ray image derived from EUV observations (Su et al. 2018), such as the observations of the Atmospheric Imaging Assembly (AIA; Lemen et al. 2012), as the reference image (ground truth).

We select a RHESSI flare event which occurred on 2010 December 2 at the location near [−820″, 480″] (Figure 8(a)). We only focus on the soft X-rays to reduce the possible contribution from the non-thermal components. Figures 8(b), (c) show the 4–10 keV image calculated from the differential emission measure (DEM) maps (e.g., Cheung et al. 2015; Su et al. 2018; Li et al. 2022) that are derived from the AIA images (henceforth DEM-X), and the reconstructed RHESSI CLEAN map (using detectors 3–6, 8, 9F) in the same energy range, respectively. The calculated DEM-X source may not be a fully

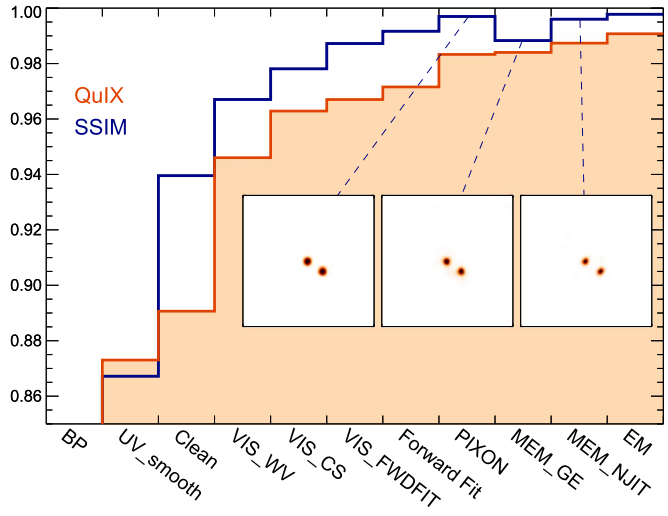

Figure 7. Comparison of the QuIX scores and SSIM scores of different imaging algorithms. The three inset images are reconstructed by using the algorithms corresponding to where the dashed line connects. Note that the scores of BP are too small (<0.6) to show in the same graphic coordinate range.

precise estimation of the real source, because the observed photons of 4–10 keV may contain contribution from non-thermal emissions in some cases (e.g., Li et al. 2022), but DEM can only represent thermal plasma. Besides, there are limitations in DEM calculations (Su et al. 2018), making it difficult to get accurate DEM maps. However, it does provide us a realistic source as the simulation ground truth, which is usually absent in near practical cases.

The simulation results, i.e., the reconstructed images by the 11 algorithms utilizing the RHESSI simulation routine, are displayed in Figure 8. The absolute QuIX score is calculated for each imaging result, and marked in each panel of Figure 8(d) in descending order. It is predictable that the BP image displays the worst result because of the heavy sidelobes straightly generated by linear BP processes (Hurford et al. 2002). The highest value of QuIX is from the VIS_FWDFIT image, which is reconstructed by fitting the function of a loop source with several parameters. However, the VIS_FWDFIT result may largely depend on the initial parameters of the source(s). In addition, there are other problems of the fitting methods in both Forward Fit and VIS_FWDFIT. More DEM-X tests can be found in Yu et al. (2025). The second best imaging result is from MEM_NJIT, and it is visually indeed the closest image to the DEM-X source. Overall, the QuIX scores can provide a proper assessment of the performance of the imaging algorithms.

It should be pointed out that we did not explore more imaging results using different parameter settings of these algorithms, which may change source sizes/width/shape and therefore affect QuIX values. In the next section, we present the effects of imaging parameters on image quality which can be revealed by the QuIX test. However, the detailed study of these parameters is beyond the scope of this work.

To show the difference between QuIX and one of its components, SSIM, we compared QuIX with SSIM values for the imaging results from the 11 algorithms (Figure 9). The imaging algorithms are sorted in ascending order based on the QuIX values, while the corresponding SSIM values do not show a monotonic trend, whereas the Forward Fit, VIS_WV, and MEM_GE algorithms show decreasing SSIM values as their QuIX increases. By visually inspecting the imaging results of these three algorithms, we think that the QuIX values of these images are more consistent with the trend of improving quality.

5.2. Investigation of the Effect of Imaging Parameters

The choice of imaging parameters may largely affect the results. However, it has been a difficult task to determine the best settings of these parameters for different instruments. QuIX may help in such studies. Here we show the application of QuIX in determining imaging parameters by testing the effect of different settings of CLEAN beamwidth factor (BWF) and imaging field of view (FOV) on image quality.

We use the DEM-X source in Section 5.1 as simulation input and reconstruct images by CLEAN, EM, and MEM_GE algorithms with different imaging FOVs, starting from 20″. The image qualities are displayed by both QuIX and SSIM in Figure 10. The results of these three algorithms are not satisfactory when the FOV is small. The corresponding QuIX and SSIM values are also small. It is expected since we used a relatively large source in this test and an FOV smaller than the source should not result in a good image. As the FOV increases, both QuIX and SSIM rapidly reach the maximum. Although both metrics show a slight decrease with further increases in FOV, the overall diminution is minimal. One reason is that the increase in FOV introduces more pixels with zero intensity, which increases the number of bases and thus reduces the difference between images. The other reason is that all three algorithms do not produce obvious sidelobes.

The test results shown in Figure 10 further demonstrate the importance of selecting an appropriate imaging FOV. In practical X-ray imaging research, a small FOV is desired whenever possible in the expectation that finer structures can be seen without significant sidelobes, while imaging quality, source size and imaging grids constrain the minimum value of the FOV range. For the DEM-X source test, which is thin in width and moderate in length, Figure 10 can reveal the recommended imaging FOV for a source of similar size. The optimal imaging FOV is $\sim 180'' \times 180''$ for the CLEAN algorithm and $\sim 140'' \times 140''$ (i.e., ~3–4 times of the source size) for the EM and MEM_GE algorithms. Overall, after the FOV reaches $100'' \times 100''$ (which is usually related to the

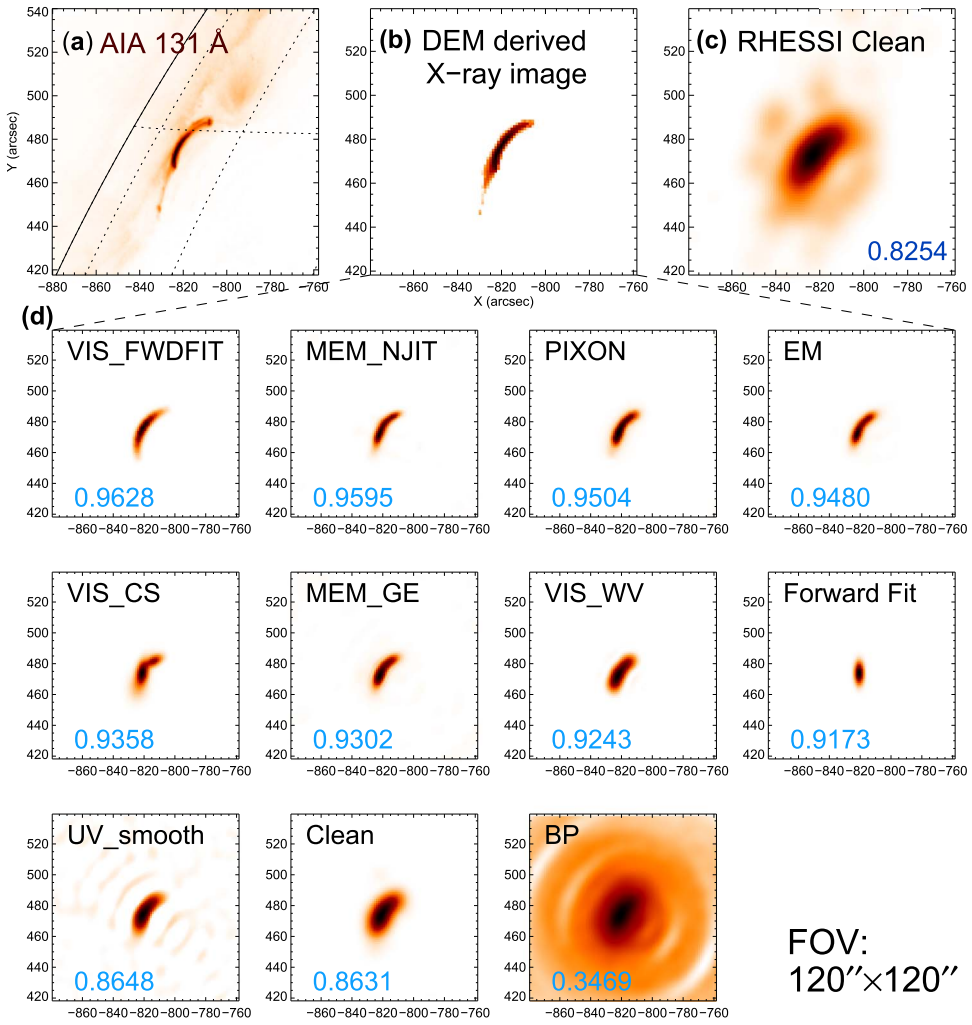

Figure 8. QuIX test result of different imaging algorithms for the DEM-based soft X-ray source. Panel (a): the AIA 131 Å map of a flare loop observed on 2010 December 2. Panel (b): the X-ray image of 4–10 keV derived from AIA DEM maps. Panel (c): the 4–10 keV map reconstructed from RHESSI X-ray data using CLEAN algorithm. Panel (d): the imaging results of the 11 imaging algorithms with a simulation input from the AIA DEM-X image. Here for CLEAN, the CBWF is 1.7.

coarsest grids used for imaging), the QuIX values for all three algorithms exhibit only slight variations.

CLEAN beamwidth factor (CBWF) is another important parameter that may directly affect the CLEAN imaging result because it determines the FWHM of the Gaussian CLEAN beam (Hurford et al. 2002) that will convolve with CLEAN components for the final image (larger CBWF corresponds to smaller FWHM). However, it is controversial to determine the optimal value of the factor and it lacks a robust approach to evaluate the cleaned image (Kontar et al. 2010; Piana et al. 2022). Figure 11 shows the application of QuIX to evaluate the cleaned images with different CBWF. Here we do not attempt to determine the best setting for CBWF as it also depends on many other factors, such as the property of the source, the combination of detectors or subcollimators, etc. In Figure 11, we only present the effect of CBWF alone, and the corresponding changes in the reconstructed images and the QuIX values. A more comprehensive test on CBWF can be found in Yu et al. (2025).

The above tests show two examples of the application of QuIX index for the investigation of two imaging parameters in X-ray image reconstruction.

5.3. Application in the Test of HXI Imaging

The HXI has a total of 91 subcollimators/detectors for imaging (Zhang et al. 2019), and several imaging algorithms have been developed for HXI (Su et al. 2019). The pitch widths of grids and the length of the HXI collimator determine the spatial resolution of the X-ray image. HXI detectors D1 to D8

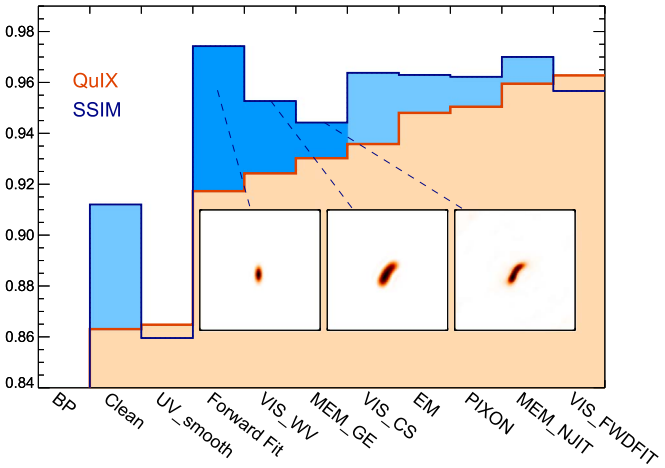

Figure 9. Comparison of the QuIX scores and SSIM scores of different imaging algorithms for the case with DEM-based X-ray source. The three inset images are reconstructed by the three corresponding algorithms. Note that the scores of BP are too small (<0.6) to show in the same graphic coordinate range.

correspond to the spatial resolution of $3\overset{''}{.}1$, D9 to D18 correspond to $4\overset{''}{.}5$, D19 to D28 correspond to $6\overset{''}{.}5$, D29 to D38 correspond to $9\overset{''}{.}3$, D39 to D48 correspond to $13\overset{''}{.}4$, D49 to D58 correspond to $19\overset{''}{.}3$, D59 to D68 correspond to $29\overset{''}{.}6$, etc. The reconstructed image changes with different combinations of imaging detectors and different imaging algorithms. The sources may be over-resolved when the ground truth sources have larger scale structures than the spatial resolution of the finest grids used for imaging. From this point, not only the selection of imaging algorithms but also the choice of detectors is noteworthy.

Here we use the simulated double Gaussian sources ($\sigma = 2''$) shown in Figure 12(b0) as the ground truth. The flux of the right source is twice that of the left one. The image size is $129'' \times 129''$. Figure 12(a) shows the QuIX values of the images from five imaging algorithms with detectors from D19 ($6\overset{''}{.}5$) to D91. The QuIX scores reflect the image qualities of Figures 12(b1)–(b5) corresponding to each algorithm. The algorithms that perform better are CLEAN, Forward Fit and VIS_CLEAN. If we change the image detectors to get different finest resolutions, as shown in Figure 12(c), the best imaging settings are CLEAN with detectors starting from D1, or VIS_CLEAN with detectors starting from D9. In this test case, the image qualities are better when the finer grids are used due to the small ($\sigma = 2''$) Gaussian sources. The detailed structures of a narrow source cannot be reconstructed if only coarse grids are used. On the other hand, if the fine grids with resolutions smaller than the source itself are used, it may cause over-resolved results. The Forward Fit results present a relatively stable trend than VIS_CLEAN and CLEAN when we change the finest detectors, because this method can get the best fit parameters of the sources when the correct number and shapes of the sources are given. Moreover, we did not change the CBWF (1.7 for default) when the detector range changes, which means that the CLEAN images can be further improved. BP and VIS_BP maps have the lowest QuIX as expected due to the heavy sidelobes.

Practically, we should roughly estimate a reasonable source size before selecting detector ranges. On the other hand, the fine grids are much easier to be affected by non-ideal factors (e.g., internal shadowing, distortions). Thus, it is unreliable to use them for imaging before the grids calibration is completed (Su et al. 2024). Currently, detectors starting from D19 or D29 are recommended for imaging in most cases.

6. Summary and Discussion

In this paper, we develop a new synthesis index of QuIX to evaluate the quality of reconstructed X-ray images. The QuIX index contains SSIM, PSNR and the percentage proximity degree PPD, which is specifically designed by considering the imaging characteristics of X-ray image reconstruction. We established an evaluation scheme of X-ray imaging quality by systematically testing a variety of imaging quality assessment indices in six aspects, and utilizing the TOPSIS method for the selected indices. An operational procedure was developed for both the new index and the overall evaluation scheme. From the tests of the imaging of DEM-based soft X-ray sources, we found that the QuIX metric is a useful tool for the selection of appropriate imaging algorithm, imaging parameters, and detectors.

To improve the reliability of QuIX, we set some changeable parameters for further optimizations. The integral limits (a and b in Equation (2)) of the percentage proximity degree PPD determine the range of pixel values taken into account. The weight of each of the three indices, RMSE, SSIM, and PPD, can be adjusted for different emphasis. As suggested in Wang et al. (2003), we also set a variable to partially utilize the SSIM index, the region of interest of imaging, which is set to be equal to the whole FOV by default. The specially designed QuIX in this work is not conclusive and definitive, and further testing and refinement are still needed. For example, we noticed in Figure 8 that the source in the FF image is clearly different from that in the DEM-X image, yet it receives a higher QuIX score than the CLEAN image. In Section 5.2, there are grainy structures within the reconstructed loop in the image for CBWF = 3.8 with the highest QuIX (which may be related to the fixed detectors used for the imaging in the test). Further optimizations can be performed to allow QuIX to focus more on the shape similarity of the reconstructed images.

The establishment of QuIX is significant for the design and optimization of X-ray imaging system (where instrument PSF of different designs can be quantitatively tested) and testing of reconstruction algorithms through imaging simulations. For the practical X-ray observational studies, the reference image of

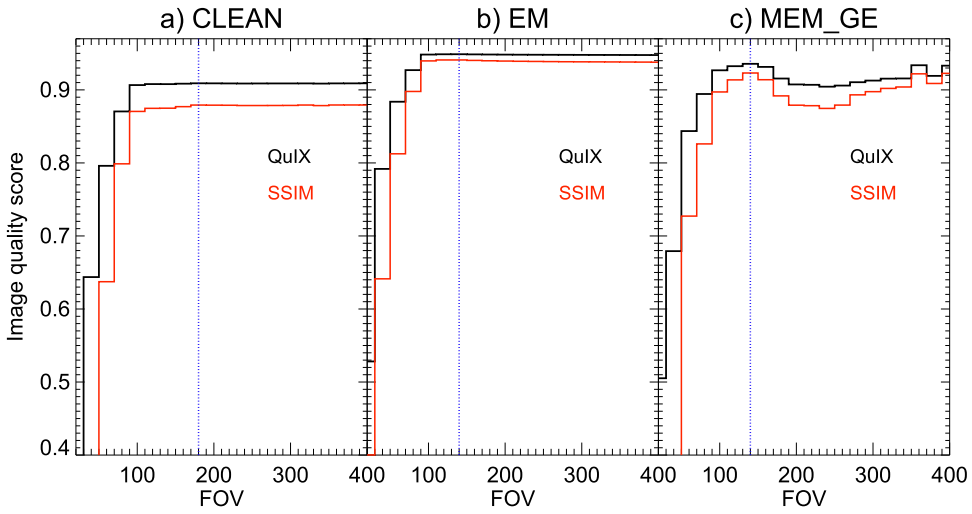

Figure 10. The QuIX and SSIM scores of three imaging algorithms, CLEAN, EM and MEM_GE, with different imaging field of views (FOVs). The imaging FOV changes from 20″ to 400″ by a 20″ interval. The vertical dotted line in each panel marks the smallest FOV of the maximum value of QuIX.

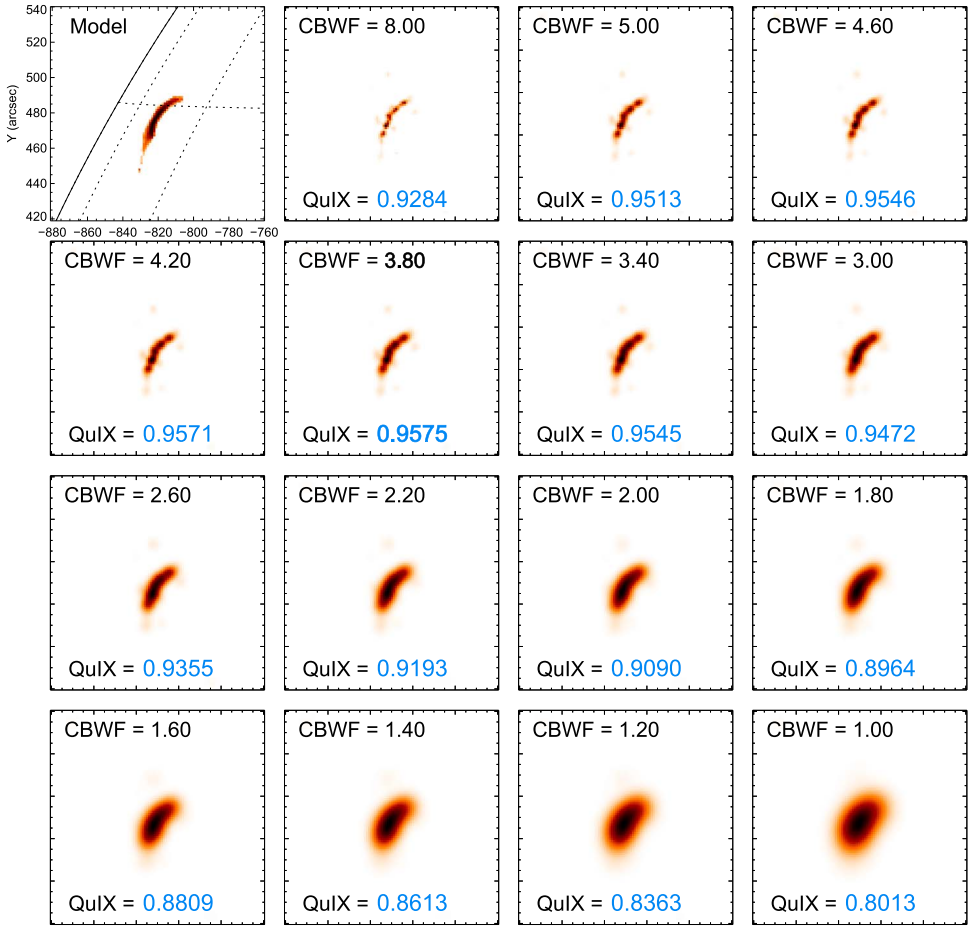

Figure 11. The test result of imaging with different CLEAN beamwidth factors. The first panel shows the DEM-based X-ray source model in Section 5.1. The rest panels show the reconstructed images of the CLEAN algorithm with decreasing `clean_beam_width_factor` from 8 to 1.

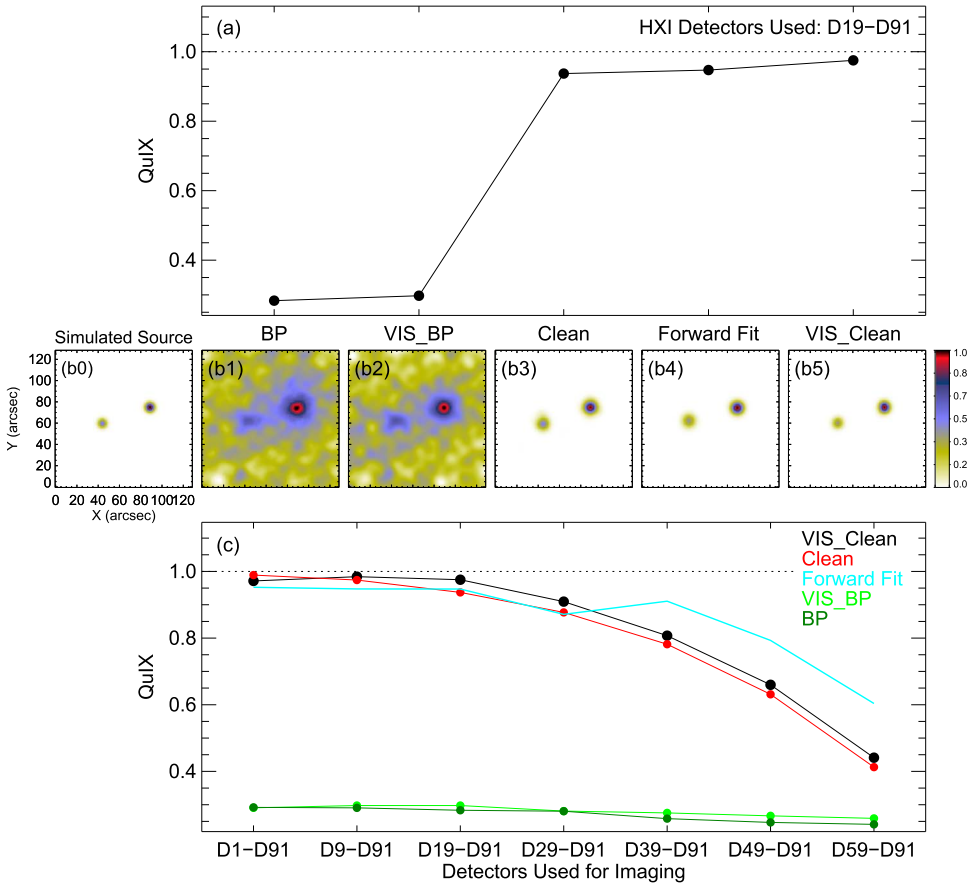

Figure 12. The QuIX results of HXI imaging with different imaging algorithms and detector ranges. Panel (a): the QuIX scores of different imaging results reconstructed by five algorithms utilized in HXI, BP, VIS_BP, CLEAN, Forward Fit, and VIS_CLEAN. The detectors used for imaging range from D19 to D91. Panel (b0): the simulated X-ray sources. Panels (b1)–(b5): the reconstructed maps by the corresponding algorithms. Panel (c): the QuIX scores of imaging results reconstructed by the five algorithms using different sets of HXI detectors.

X-ray source is usually unavailable, which severely limits the application of QuIX. In this case, we can use the methods of measuring the distortion level of the reconstructed image compared with the observed signal to evaluate the X-ray imaging quality, such as the C-statistic and reduced χ^2 (e.g., Dennis & Tolbert 2019; Massa et al. 2022), despite that they do not measure the structural similarity and the amount of sidelobes.

Acknowledgments

The work is supported by the National Natural Science Foundation of China (NSFC) 12333010, the National Key R&D Program of China 2022YFF0503002, the Strategic Priority Research Program of the Chinese Academy of Sciences (grant No. XDB0560000), and the NSFC 11921003. Z.L. is also supported by the Prominent Postdoctoral Project of Jiangsu Province (2023ZB304). ASO-S mission is supported by the Strategic Priority Research Program on Space Science, the Chinese Academy of Sciences, grant No. XDA15320000. The authors express their gratitude to the colleagues at Purple Mountain Observatory for their constructive suggestions and comments.

ORCID iDs

Zhen-Tong Li https://orcid.org/0000-0002-4230-2520
Wen-Hui Yu https://orcid.org/0009-0004-2847-9540
Yang Su https://orcid.org/0000-0002-4241-9921
Wei Chen https://orcid.org/0000-0001-5279-3266
Wei-Qun Gan https://orcid.org/0000-0001-9979-4178

References

Aschwanden, M. J., Metcalf, T. R., Krucker, S., et al. 2004, SoPh, 219, 149
Aschwanden, M. J., Schmahl, E. & RHESSI Team 2002, SoPh, 210, 193
Benvenuto, F., Schwartz, R., Piana, M., & Massone, A. M. 2013, A&A, 555, A61
Chen, W., Su, Y., Zhang, Z., et al. 2020, AcASn, 61, 41
Cheung, M. C. M., Boerner, P., Schrijver, C. J., et al. 2015, ApJ, 807, 143
Dennis, B. R., & Tolbert, A. K. 2019, ApJ, 887, 131
Duval-Poo, M. A., Piana, M., & Massone, A. M. 2018, A&A, 615, A59

Felix, S., Bolzern, R., & Battaglia, M. 2017, ApJ, 849, 10
Fletcher, L., & Hudson, H. S. 2002, SoPh, 210, 307
Gan, W., Zhu, C., Deng, Y., et al. 2023, SoPh, 298, 68
Gan, W.-Q., Zhu, C., Deng, Y.-Y., et al. 2019, RAA, 19, 156
Högbom, J. A. 1974, A&AS, 15, 417
Holman, G. D., Aschwanden, M. J., Aurass, H., et al. 2011, SSRv, 159, 107
Hudson, H., & Ryan, J. 1995, ARA&A, 33, 239
Hurford, G. J. 2010, ISSIR, 9, 223
Hurford, G. J., Schmahl, E. J., & Schwartz, R. A. 2005, AGUSM, 2005, SP21A
Hurford, G. J., Schmahl, E. J., Schwartz, R. A., et al. 2002, SoPh, 210, 61
Hwang, C.-L., & Yoon, K. 1981, Methods for Multiple Attribute Decision Making, Multiple Attribute Decision Making: Methods and Applications A State-of-the-art Survey (Berlin: Springer), 58
Islam, S. M. R., Huang, X., & Le, K. 2013, in Neural Information Processing, ed. M. Lee et al. (Berlin: Springer), 549
Jagalingam, P., & Hegde, A. V. 2015, AqPro, 4, 133
Kellermann, K. I., & Moran, J. M. 2001, ARA&A, 39, 457
Kong, X., Chen, B., Guo, F., et al. 2022, ApJL, 941, L22
Kontar, E. P., Hannah, I. G., Jeffrey, N. L. S., & Battaglia, M. 2010, ApJ, 717, 250
Kosugi, T., Makishima, K., Murakami, T., et al. 1991, SoPh, 136, 17
Krucker, S., Hurford, G. J., Grimm, O., et al. 2020, A&A, 642, A15
Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, SoPh, 275, 17
Li, Z., Su, Y., Veronig, A. M., et al. 2022, ApJ, 930, 147
Lin, R. P., Dennis, B. R., Hurford, G. J., et al. 2002, SoPh, 210, 3
Liu, Z., Blasch, E., Xue, Z., et al. 2012, TPAMI, 34, 94
Massa, P., Battaglia, A. F., Volpara, A., et al. 2022, SoPh, 297, 93
Massa, P., Schwartz, R., Tolbert, A. K., et al. 2020, ApJ, 894, 46
Massone, A. M., Emslie, A. G., Hurford, G. J., et al. 2009, ApJ, 703, 2004
Metcalf, T. R., Hudson, H. S., Kosugi, T., Puetter, R. C., & Pina, R. K. 1996, ApJ, 466, 585
Ogawara, Y., Takano, T., Kato, T., et al. 1991, SoPh, 136, 1
Piana, M., Emslie, A. G., Massone, A. M., & Dennis, B. R. 2022, Hard X-Ray Imaging of Solar Flares (Berlin: Springer)
Puetter, R. C. 1995, IJIST, 6, 314
Saint-Hilaire, P., Krucker, S., & Lin, R. P. 2008, SoPh, 250, 53
Sara, U., Akter, M., & Uddin, M. S. 2019, Journal of Computer and Communications, 7, 8
Schmahl, E. J., Pernak, R. L., Hurford, G. J., Lee, J., & Bong, S. 2007, SoPh, 240, 241
Schwarz, U. J. 1978, A&A, 65, 345
Sheikh, H., & Bovik, A. 2006, ITIP, 15, 430
Sheikh, H., Bovik, A., & de Veciana, G. 2005, ITIP, 14, 2117
Su, Y., Liu, W., Li, Y.-P., et al. 2019, RAA, 19, 163
Su, Y., Veronig, A. M., Hannah, I. G., et al. 2018, ApJL, 856, L17
Su, Y., Zhang, Z., Chen, W., et al. 2024, SoPh, 299, 153
Wang, Z., & Li, Q. 2011, ITIP, 20, 1185
Wang, Z., Simoncelli, E., & Bovik, A. 2003, in The Thrity-Seventh Asilomar Conf. on Signals, Systems Computers, 2003, Vol. 2, 1398
Yu, W., Su, Y., Li, Z., Chen, W., & Gan, W. 2025, RAA, in press
Zhang, L., Zhang, L., Mou, X., & Zhang, D. 2011, ITIP, 20, 2378
Zhang, X., Ye, P., & Xiao, G. 2020, in 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), 468
Zhang, Z., Chen, D.-Y., Wu, J., et al. 2019, RAA, 19, 160