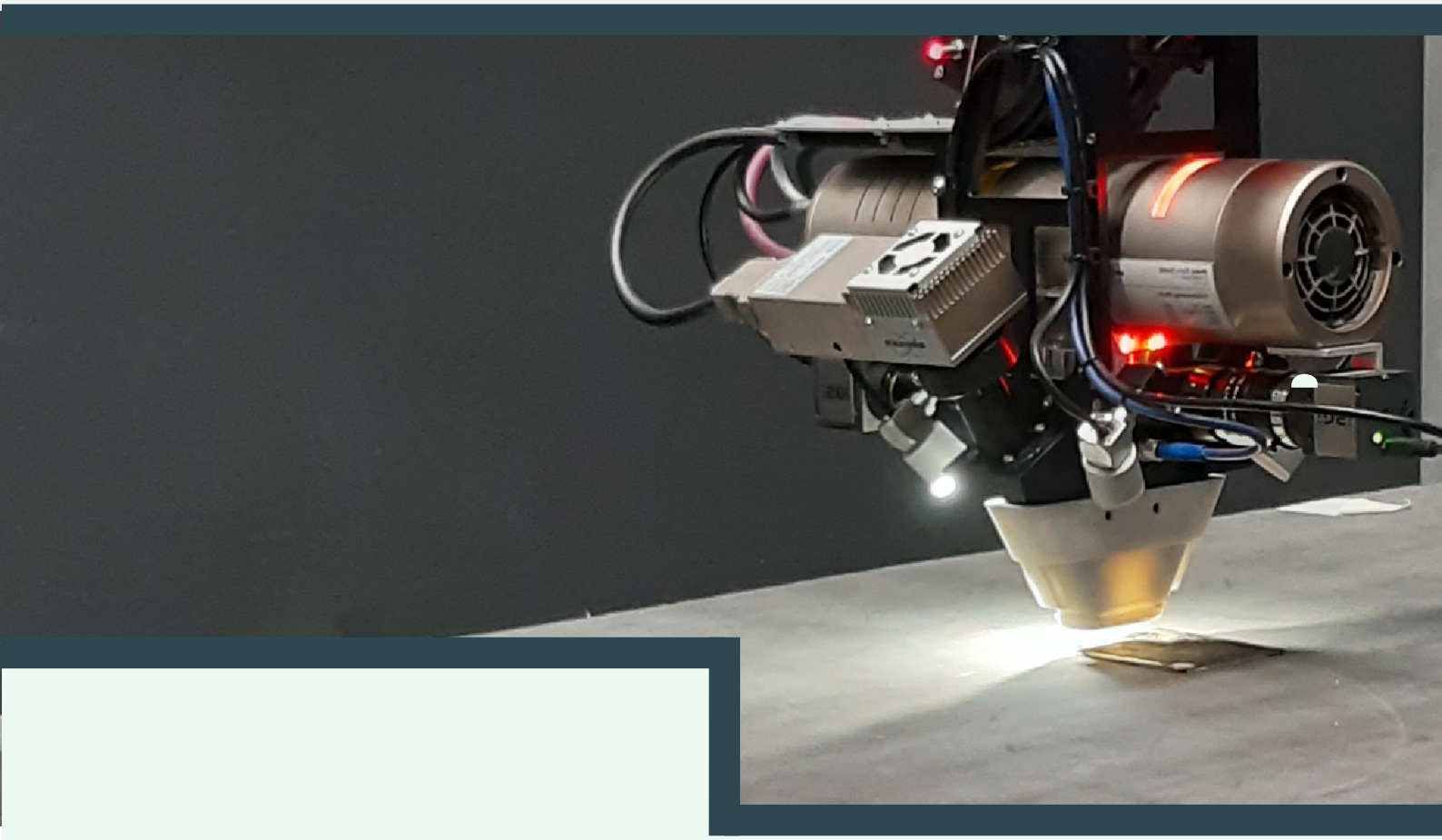
A Thesis by Maartje Huijbrechts

October 23, 2023



Accelerating MA-XRF Data Acquisition by Exploiting Local Spatial and Spectral Relations within a Hyperspectral Datacube

An Approach through Wavelet Denoising

TU Delft

Abstract: Macro X-ray fluorescence (MA-XRF) is a recently developed technology allowing to obtain elemental information from cultural heritage objects. This information can, for example, be used to identify pigments used in a painting. Yet, the extended period of time it takes to scan an object is a major issue within MA-XRf. For instance, it took about 60 days to scan the Ghent Altarpiece. The long scanning time is a consequence of the necessary dwell time per pixel to create a robustly interpretable spectrum: the higher the dwell time, the higher the signal-to-noise ratio (SNR), hence, the easier to detect elements. This thesis explores a possible solution for this problem using a denoising algorithm that increases the signal-to-noise ratio post-acquisition by exploiting the similarity between neighbouring pixels and spectra. To this end, a customized method of wavelet filter bank denoising is proposed. Current thresholding methods used in wavelet filter bank denoising are not suitable for filtering MA-XRF data, therefore, a novel thresholding method is introduced. Here, the widely used universal thresholding method is used as a basis, for which the formula for calculating the standard deviation of the detail coefficients of a channel is altered. Several design parameters of wavelet filter bank denoising were evaluated using a synthetic dataset, for which the performance quality indicators root mean square error (RMSE), mean absolute error (MAE) and SNR were determined. The parameters for which we optimized were the *mother wavelet*, the *number of decomposition levels*, and the *number of neighbouring channels used for determining the standard deviation σ for thresholding*. Good performance was obtained with the haar, db2, and coif1 wavelets, all at 3 levels of decomposition. A suitable number of neighbouring channels depended on the decomposition level and was determined to be 3 (on each side of the channel). Herewith, the signal-to-noise ratio was improved for both the average pixel spectra and the sum spectrum. The filtered synthetic dataset simulated to have a dwell time of 0.5 seconds had a SNR approximately equal to the raw synthetic dataset simulated to have a dwell time of 0.75 seconds. Hence, the algorithm succeeded in lowering the necessary dwell time. A case study of a daguerreotype was used to test the proposed denoising algorithm.

Contents

1 Introduction

Investigating art objects and archaeological artefacts dates back to the fifteenth century.[1] More than a century ago, this endeavour was extended by investigating these objects with technological means.[2] Cultural heritage objects were researched with a variety of technologies, such as X-ray radiography, ultraviolet fluorescence, and infrared radiography. In 2007, a new technique, known as macroscopic X-ray fluorescence (MA-XRF), was developed specifically for the research of art.[3] With this novel and non-destructive method of investigation, elements present in a sample can be detected over a large surface,[4] instead of just one point, as was already possible using X-ray fluorescence (XRF). Herewith, elemental distribution images can be made, which, when compared to a visual light image of the sample, can result in the identification of pigments used within the sample.[5]

A problem faced in MA-XRF is that scanning a sample often takes up a lot of time. For example, in 2016, it took 60 days to scan *The Ghent Altarpiece* by the van Eyck brothers, which covered 8 m^2.[6] The limited amount of time a cultural heritage object is available for research, but also the desire to minimize the exposure of an object to X-rays, creates the demand to reduce the amount of necessary scanning time.[7]

The reason that scanning takes up so much time is the fact that, during MA-XRF, each pixel on a painting is scanned individually; taking up tenths of a second per pixel.[7] When more than 16 million pixels have to be scanned, as was the case with *The Ghent Altarpiece*,[8] the scanning time per pixel a.k.a. the dwell time, becomes significant.

One way to reduce the overall scanning time is to minimize the dwell time. However, low dwell times result in low signal-to-noise ratios (SNR) and are, therefore, accompanied by problems with accurate signal interpretation. In other words, there is a trade-off between dwell time and the detectability of elements. This research aims to reduce the dwell time, without compromising the detectability by increasing the spectral signal-to-noise ratio of macroscopic X-ray fluorescence data after it is collected but before it is evaluated.

At the moment, there are several methods used for processing MA-XRF data. Still, most procedures that are applied do not address the issue of denoising to improve a low SNR. The processing methods focus, above all, on fitting the individual pixel spectra to find the sample's elemental composition. This seems to be a consequence of the fact that MA-XRF has been developed from XRF. In the latter, the SNR was not much of an issue, since an XRF measurement could easily take up several seconds for it was merely one point that had to be measured. Chopp *et al.*[9] mention the problem of limited denoising being done in the field of MA-XRF. Moreover, besides stating the problem, they gave the recommendation to apply image-denoising methods to denoise the MA-XRF data. Therefore, the

aim of this research, which is to increase the spectral signal-to-noise ratio, will be attained by exploiting both spatial relations between neighbouring pixels and spectral relations between neighbouring channels. For this purpose, a customized wavelet filter bank denoising method is proposed.

This thesis is built up as follows: As an extension of this brief introduction, the working principle of MA-XRF and the data characteristics are explained. Then, the state-of-the-art methods for noise reduction and data fitting are discussed, resulting in the identification of the research gap and a research proposal. Hereafter, an explanation of the proposed denoising method, wavelet filter bank denoising, is given; including a description of a customized thresholding method. Furthermore, a qualitative analysis of MA-XRF data is performed to clarify how the proposed denoising method can be applied. Then, the method for testing the filter designs and a case study of a daguerreotype is discussed. This is followed by the results and discussion of the experiments and the case study. The thesis is finalised with the conclusion where we summarize the findings on whether the proposed denoising method improves the signal-to-noise ratio and in effect lowers the necessary dwell time for MA-XRF scanning.

1.1 MA-XRF

MA-XRF owes its name to the fact that large areas are being scanned using XRF. The surface is scanned pixel by pixel, obtaining thousands of point spectra. In this section, first the working principle of XRF, from which MA-XRF is developed,[5] will be explained after which a brief explanation of the resulting spectra follows. Subsequently, MA-XRF and its possibilities will be discussed.

1.1.1 X-ray Fluorescence (XRF)

XRF is a non-destructive method that provides insight into the elemental composition of a sample. During a measurement, X-rays are emitted by an X-ray tube after which they hit the material. Atoms present in the material become excited and fluoresce X-ray photons, which can be detected by a silicon drift detector (SDD). The energies of the photons are characteristic of the elements they originate from and, therefore, signify elements present in the sample. The process of creating fluorescing X-ray radiation is visualised at the atomic level in Figure 1. When X-rays are incident on the atom, electrons, which absorb the energy of the X-ray photons, are ejected from their shell; this action is shown by the red arrows. The ejected electron can be neglected for data-evaluative purposes since the electron is absorbed either within the sample or in the air between the sample and the detector due to its mean free path in air being too short to reach the SDD.[10] However, through ejecting the electron, the atom has become ionized. The atom

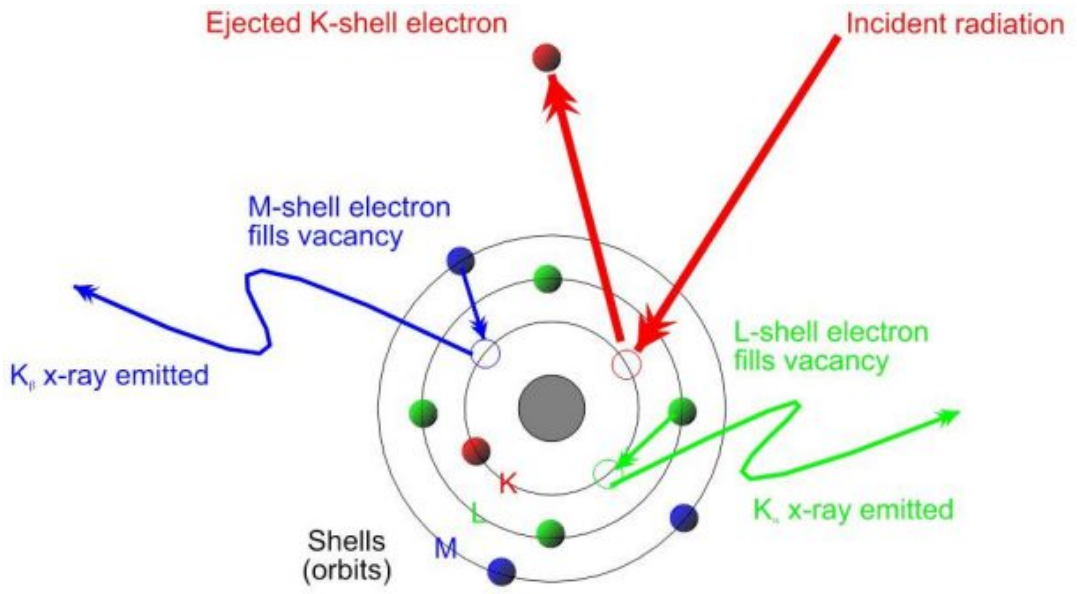


Figure 1: Characteristic X-ray production.[11]

seeks to be in equilibrium, and therefore, electrons from outer shells fall back to fill the vacancies in the inner energy shells. During this process, the energy difference between the outer and the inner shells is emitted in the form of X-rays. These emitted X-rays, visualised in blue and green, are the characteristic energies that are measured.

1.1.2 Spectrum

X-rays emitted from the sample are detected by an SDD and, subsequently, processed in order to obtain a spectrum. An X-ray spectrum is usually plotted having the intensity (photon counts) on the vertical axis and energy channels in keV on the horizontal axis. An example of a spectrum is given in Figure 2. The characteristic energy that originates from atoms within the sample hits the SDD and ionizes atoms in the semi-conductor crystal. The resulting charge is then converted to a voltage of similar proportions after which the voltage is binned by a pulse processor, obtaining the spectrum.[12] The more photons of a certain energy are detected, the higher the peak. In the spectrum in Figure 2, several peaks are

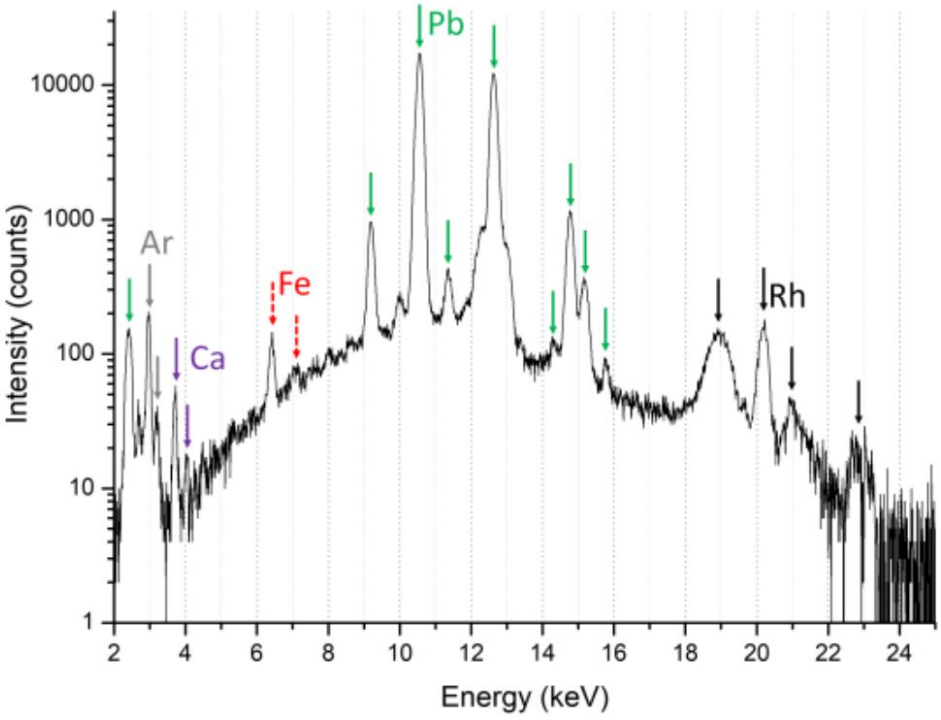


Figure 2: Sum spectrum extracted from the forehead of the virgin from the panelled vault in the church of Le Quillio.[13]

identified as being the characteristic radiation for certain elements such as lead (Pb) and calcium (Ca). In this manner, a spectrum can be analyzed and the elemental composition of the sample determined.

1.1.3 Macroscopic X-Ray Fluorescence (MA-XRF)

As was already mentioned in the introduction, MA-XRF is a technique for scanning an entire surface instead of just a point. The surface is scanned pixel by pixel, adopting a stepsize of typically 100-1000 μm.[14]. From every pixel a spectrum is obtained resulting in a hyperspectral datacube, visualised in Figure 3. The x and y axes represent the spatial pixel position, while the z-axis represents the energy channels in which the photon counts are stored per pixel, in other words, the spectrum. The brown bar represents how the photon counts per channel, obtained for every pixel, are stored in the datacube. On the side of the z-axis, the spectrum of that pixel is visualised. The advantage of a MA-XRF scan over an

XRF measurement is dual. For example, the spectra can be visualised together, as a sum of individual spectra, called a sum spectrum, as is done in Figure 2. By summing spectra from a specific part of a painting, for example, where the colour is identical, the elemental composition of that part, and as such, possible pigments used, could be more accurately determined. Moreover, looking at an entire region instead of just one point helps reduce the effect of noise. Besides looking at a specific region to find its elemental composition — the spectrum on the z-axis — it is also possible to visualise where in the sample a specific element is present and in which concentration, which, in this research, will be referred to as peak intensity maps. A peak intensity map is generated by summing channel maps covering an elemental peak. Such a channel map is visualised on the x, y axes as the beige rectangle in Figure 3. The peak intensity maps are grey-level images in which pixels with a higher photon count in a specific channel are depicted lighter and pixels with a lower photon count are shown darker.

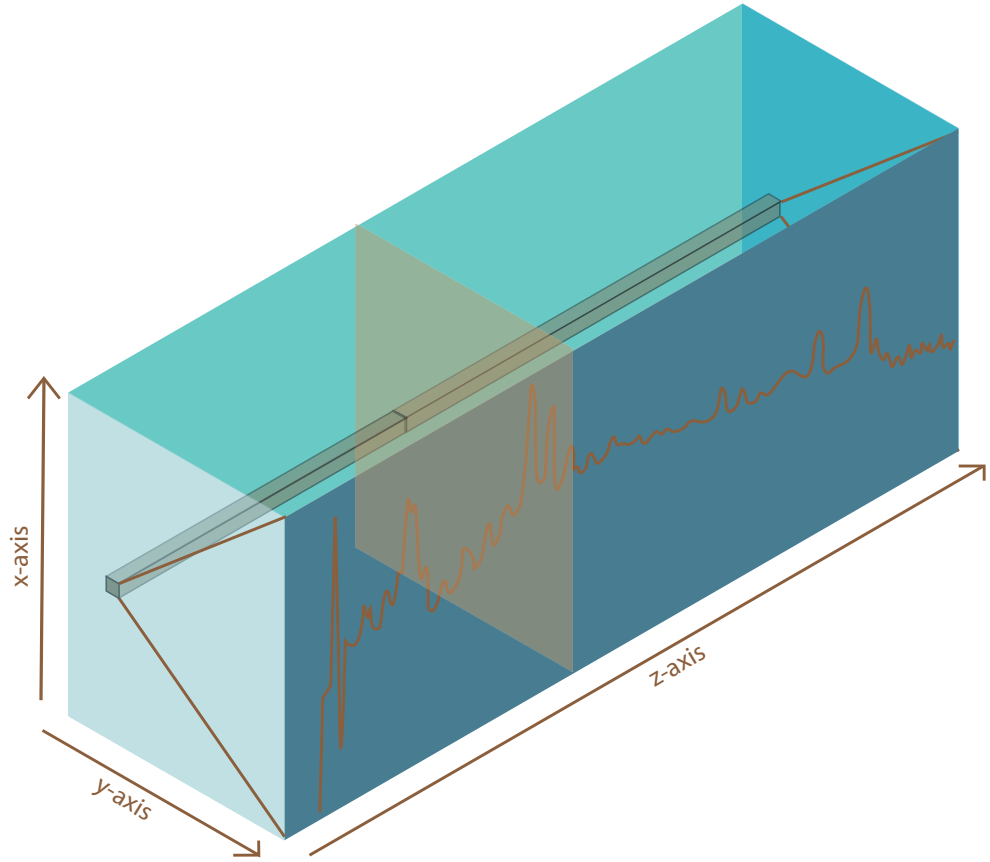


Figure 3: Schematic overview of a hyperspectral datacube. The x and y axes represent the spatial pixel position, while the z-axis represents the energy channels. The brown beam represents the spectrum of one pixel, this spectrum is visualised on the side of the cube. The beige rectangle represents the pixel values of all the pixels in one channel.

1.2 MA-XRF Data

The hyperspectral datacube that is obtained during scanning does not instantly provide us with knowledge of elemental compositions of certain regions or elemental distribution images; the spectrum has to be interpreted first. There are, however, aspects that complicate the interpretation of these spectra: the signal of the characteristic radiation is disturbed by noise.

The subject of data acquisition was discussed in a simplified manner in subsection 1.1. However, it is necessary to consider the procedure in more detail in order to understand the problems faced during data-evaluation. In this chapter, an overview will be given of the major factors that contribute to the spectrum, first discussing the characteristic radiation, after which different kinds of noise are addressed. Finally, the detection limit is discussed.

1.2.1 Characteristic Radiation

In section 1.1.1 it was explained how incident X-rays resulted in the emittance of X-ray photons from elements in the sample. In this section, we explain more in-depth how the characteristic energy patterns are obtained.

The amount of energy that is emitted by the atom depends on the energy differences between the electron shells from which the electrons are ejected and the shells that supply the electrons that fill the vacancies; these differ per element. When regarding the schematic depiction of an atom in Figure 1, the three shells closest to the nucleus have been depicted, called the K, L, and M shells. Next, the straight green and blue arrows represent electrons from an outer shell filling the vacancy in the inner shell. When an electron from the L shell fills the vacancy in the K shell (green arrow), it is called K_α radiation and when an electron from the M shell fills the vacancy in the K shell (blue arrow) it is called K_β radiation. Also, electrons from the L shell and M shell can be ejected from their shell and filled by electrons from outer shells, which is indicated with similar nomenclature as can be seen in Figure 4.

Besides the photon-energies, the relative photon counts are characteristic as well. In the iron spectrum of Figure 4, it can be seen that the K_α peak is much larger than the K_β peak. The difference in counts has to do with the fact that some electrons in certain shells are more prone to becoming ejected than others. Moreover, some vacancy transfers (Coster-Kronig transitions) are more likely to happen. The subsequent emission probabilities per element are constant and properly quantified. In the end, the peak pattern of an element consists of peaks at characteristic energies, having alternating but proportional photon counts with respect to one another.

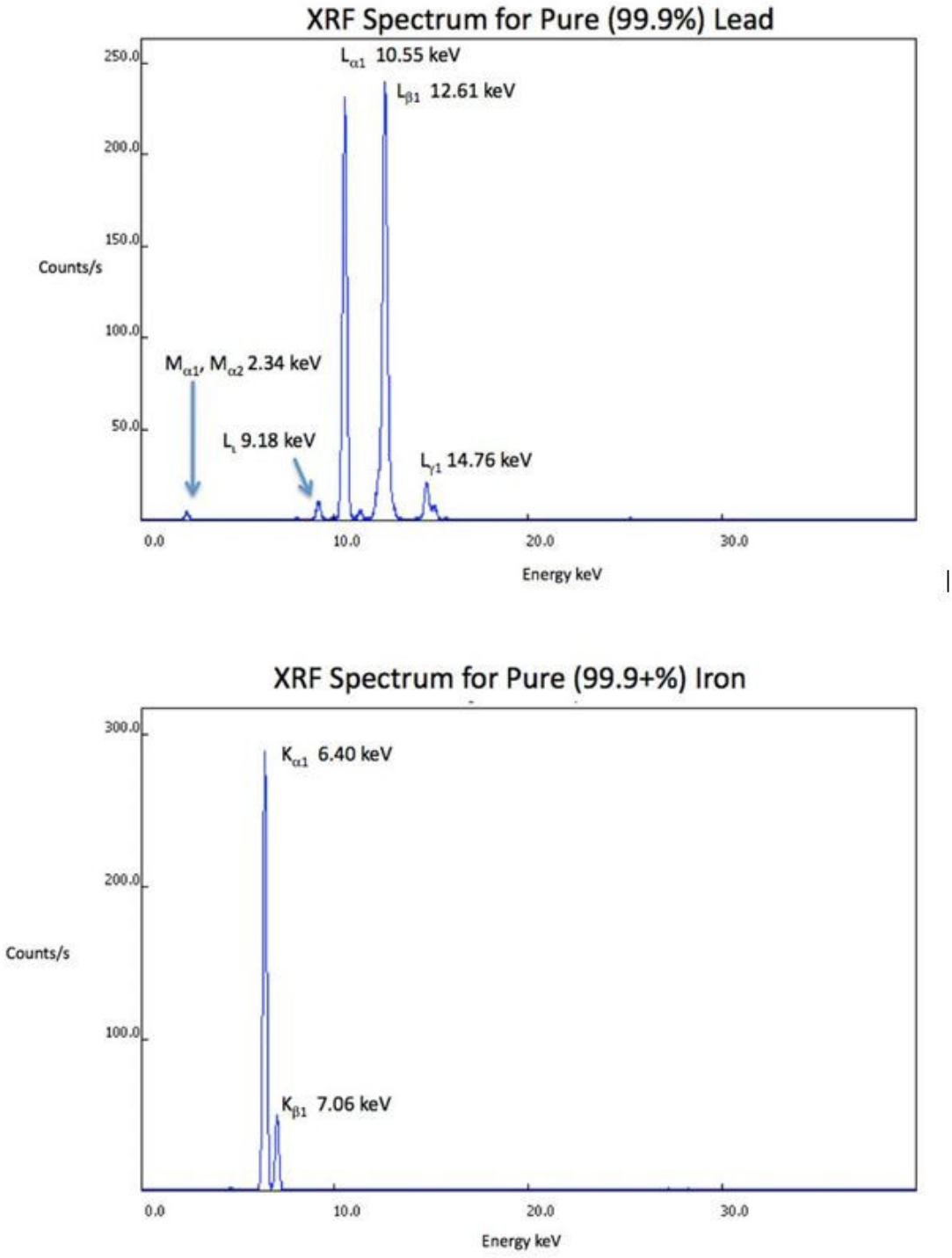


Figure 4: Spectra of the Characteristic Radiation of Lead(Pb) and Iron(Fe).[15]

1.2.2 Noise

The data that is obtained is not a perfect accumulation of elemental spectra coming from the sample; the data is perturbed by various kinds of noise. The MA-XRF scanner itself, for example, influences the data. Aspects such as detector flaws, dwell time variations, and background radiation all contribute to unwanted variations within the spectrum. Also, absorption can change the expected radiation pattern of the characteristic radiation or cause elements not to be detected at all. Finally, shot noise, which is inherent to photon-imaging, results in intensity variations following a Poisson distribution.

Background Radiation. When looking at Figure 2, a volume underneath the peaks can be perceived. This is the scattered bremsstrahlung, which is made up of photons created by electrons being decelerated by the anode in the X-ray tube. Another feature of the spectrum that also originates in the X-ray tube is a peak from the anode material; in the case of the Bruker M6-jetstream, this is a rhodium peak.[16] Elastically and inelastically scattered bremsstrahlung and characteristic rhodium radiation form the background radiation.

Dwell Time Variations. Dwell time variations are unintended variations in the time that a pixel is being scanned. When a pixel is scanned for a longer time than another pixel, the amount of photons that are detected can be higher for one pixel than for the other. If only one pixel would be analysed, a difference in photon counts does not matter significantly since the proportions of characteristic line intensities remain intact. However, when two pixels are compared over a certain characteristic line, it may look like the concentration of an element in the pixel with a longer dwell time is higher than the concentration of that same element in the other pixel. Hence, dwell time variations cause photon counts to not be comparable and hence obstruct the visualisation of elemental distribution images.

Detector Flaws. The detector electronics are not capable of recording more than one event at once. Hence, when the detector is measuring an event, it becomes insensitive for a short amount of time, called dead time. Therefore, photons that are detected too close to one another are rejected by the detector electronics. Sometimes, however, when the time window is too narrow between the two (around 100 ns), the photons can be detected as a single event, hereby creating a pile-up peak in the spectrum. These peaks are equal to the summed energy of the characteristic radiation of the individual photons. The higher the count rate, the bigger the issue of pile-up peaks.[9] It is also possible that an X-ray photon coming from a silicon atom — created when incident X-rays ionize the silicon atoms in the SDD — escapes the detector measurement. The measured charge at

the anode is then the energy of characteristic radiation minus the energy of the silicon photon (1.74 keV), resulting in an escape peak.[17] Another issue with the SDD is incomplete charge collection; when not all electrons that are generated by an incoming X-ray reach the anode, the FET measures lower energy than that of the X-ray, this results in tailing in the spectrum. For low energy X-rays this problem is more significant since they have a small penetration depth in the SDD and therefore already a poor charge collection at the front contact.[12] Incomplete charge collection also results in the so-called zero strobe peak or zero peak, which appears at the zero-energy channel in the spectrum.[17] This zero peak is often advantageously used for normalizing and calibrating the spectrum.

Absorption. In the process of creating fluorescence radiation, an incident X-ray creates a series of events that results in another X-ray becoming emitted from the atom. Naturally, this emitted X-ray could also excite other atoms in the sample instead of becoming detected by the SDD. When this happens, it is called absorption. Due to the absorption of X-rays in the sample, elemental ratios of fluorescence lines, for example the K_α and K_β, change. This complicates the interpretation of the spectra, as the ratios between lines are characteristic for each element. Moreover, due to the fact that the investigated radiation has low energy - less than 30 keV - covering layers consisting of heavy elements could even prevent the detection of lighter elements in sub-surface layers.[18]

Absorption of photons does not only take place in the sample itself. Since, MA-XRF is not performed in a vacuum, X-rays are absorbed in the air between the painting and the detector as well. Hence, not all elements can be detected. This is, however, mainly a problem for lower excitation/fluorescence energies and, therefore, light elements cannot be detected outside a vacuum.[17]

Photon Shot Noise. Shot noise is inherent to imaging, being a result of the discrete nature of light over time. When scanning a pixel, photons are emitted at random, resulting in a non-uniform spread of photon counts in time, which can be modelled as a Poisson process. When scanning a pixel for a long time, this discontinuity averages out and the detected photons accurately represent the material composition of the pixel. However, when a pixel is scanned for a shorter time, the random intensity fluctuations start to dominate the signal, making it difficult to interpret the data.

1.2.3 Limits of Detection

It was mentioned in 1.2.2 that low-energy X-rays are absorbed in the air. As such, elements with a low atomic number are not detected outside a vacuum. Meanwhile, there is also an upper limit to the elements that can be detected due

to the maximum voltage of 50 kV within the X-ray tube. The limit of detection for element i is described by the following equation:[16]

$$LOD_i = 3 \times \frac{\sqrt{N_{back}}}{N_{signal}} \times c_i \times \sqrt{t} \quad (1)$$

Here, N_{signal} is the net intensity (the area of the peak minus the background) of element i, N_{back} is the intensity of the background below the peak that is analysed, c_i is the mass percentage of element i, and t is the dwell time. The lower the limit of detection, the better low-abundant signals can be discerned. *Note that c_i and $\sqrt{t}$ are normalization factors for, naturally, a longer dwell time and higher mass percentage would result in improved detectability.*

Alfeld *et al.* determined the range of detectable elements by their K-lines. The Bruker M6 Jetstream was able to detect elements between Ti (Z=22) and Mo (Z=42), for which they defined the detectability as 100 ppm for measurements of 1 second.[16] However, there is no clear limit in terms of lowest or highest detectable element since it depends highly on the weight fraction of the element, the noise, and the dwell time. It depends, moreover, on the hardware. For example, the choice of having a polycapillary lens to define the beam that is excited from the X-ray tube, results in a lower transmission for high energy X-rays originating from the tube.

The importance of the dwell time for the detectability of an element becomes apparent from Equation 1. Hence, although the mechanics of a Bruker M6 Jetstream allow for a pixel to be scanned in 1 ms,[19] an accurate spectrum can only be obtained when a pixel is scanned with a longer dwell time. Still, as was mentioned before, the dwell time cannot be too high, for then the scanning of a painting would take unpractically long. Therefore, an optimum between scanning time and proper limits of detection needs to be chosen. Past research with the Bruker M6 Jetstream used dwell times between 70 and 90 ms/pixel.[20] [21] [22] [23] However, the dwell time can differ a lot per investigation. Sometimes, smaller areas are scanned to obtain more detail; scanning each pixel for a few seconds. For example, in the case of the research of Caravaggio's painting Supper at Emmaus, Alfeld and Janssens decided to apply a longer dwell time (0.42 s) to certain areas for visualizing the elements K and Mn which were more difficult to detect. Meanwhile, areas in which there were elements for which the MA-XRF scanner was more sensitive or areas in which elements were abundantly present, were scanned with lower dwell times (below 10 ms).[24]

2 Data Processing

In subsection 1.2 the aspects that contribute to the spectrum were shown. Not only did it become apparent that not every element in the sample can be detected, but MA-XRF is also more sensitive to certain elements than to others. In combination with the present noise, the process of data evaluation demands scrutiny. The evaluation of obtained data allows us to transform the hyperspectral datacube into meaningful elemental distribution maps and/or elemental compositions of individual pixels that give insight into the material composition of the object. Currently, the process of data evaluation is often done using the software PyMCA and/or Datamuncher. In this chapter, methods for data evaluation will be discussed, starting with explaining the difference between pre-processing, processing, and post-processing. Hereafter, the research gap, considering pre-processing methods, will be stated. This chapter concludes with my research proposal.

2.1 Pre-processing, Processing and Post-processing

The methods discussed in this section can be divided into pre-processing, processing, and post-processing methods. The distinction is important to make, for it will explain the research gap that is investigated in this thesis. During data processing, the data is transformed from photon counts per energy channel into information on relative amounts of elements/compounds within the sample. Preliminary to processing, the data can be made more reliable for processing. Increasing the SNR, normalizing the data or correcting for dwell time variations can be referred to as pre-processing. During pre-processing no interpretation of the data is done, merely the signal is changed. Post-processing is done after the data has already been fitted in order to correct measurement issues or make the data more legible.

First, processing methods will be discussed. The main processing methods that are currently used for MA-XRF are iterative least squares fitting and matrix factorization, which will be discussed in section 2.1.1 and section 2.1.2, respectively. Subsequently, examples of pre-processing methods currently used in MA-XRF are given in section 2.1.3. Finally, post-processing methods are treated in section 2.1.5. Note that often a single method can be used for pre-processing as well as processing and post-processing. For example, background correction is either done before fitting the spectra or during the fitting of the spectra; hence, it can be regarded as a pre-processing or processing step. Moroever, the clustering methods that are used for processing purposes can be applied to denoise MA-XRF data as well, as is explained in section 2.1.2.

2.1.1 Iterative Least Squares Fitting

Processing

Most data-evaluation procedures that are applied to MA-XRF data use a form of least squares fitting. Herein, the measured spectrum S is approximated mathematically by a function f and the difference between the two is minimized. This process can be described with the following formula:

$$\chi_r^2 = \frac{1}{n-m} \sum_{i=0}^{n} w_i (S_i - f_i(a_0, a_1, ..., p_0, p_1, ...))^2 \tag{2}$$

Where χ_r^2 is the reduced weighted squared difference, $n - m$ signify the degrees of freedom where n is the number of channels and m the number of parameters. The variable $w_i = 1/f_i$ is the weight added per channel that ensures that elements with a high intensity do not dominate the fit and a and p are the linear and nonlinear parameters respectively that ensure χ_r^2 is minimized. Note that once nonlinear parameters are involved we are talking about a Non-Linear Least Squares fitting (NL-LS). NL-LS is the processing method adopted in the PyMCA software.[25]

Alfeld and Janssens noted two limitations of NL-LS: one being the signal-to-noise ratio not being good enough to improve the nonlinear parameters p, the other being that the procedure takes too long due to its iterative nature. They remark that the former can be dealt with by starting the fit of each individual spectrum with the same non-linear parameters p to prevent the drifting of these parameters during processing. The latter could be solved by optimizing for linear variables only, hence finding χ_r^2 in a Linear Least Squares (L-LS) procedure. In this case, the values for p are determined beforehand by fitting a spectrum that is representative for the entire set, most commonly the sum spectrum. The parameters p remain fixed during L-LS.[24]

2.1.2 Matrix Factorization

Processing

Matrix factorization methods are much faster in fitting the data than the previously discussed iterative least squares procedures. Here the measured hyperspectral datacube $C \in \mathbb{Z}^{H \times W \times C}$ is approximated by a matrix multiplication of two matrices $A \in \mathbb{R}^{H \times W \times M}$ and $F \in \mathbb{R}^{M \times C}$:

$$C \approx AF \tag{3}$$

where F is the feature matrix and A the abundance matrix. The AF approximation of the datacube C is in most cases optimized using a least squares fit.[7] There are various kinds of optimisation procedures that use matrix factorization.

In this section I will touch upon a few that are used in MA-XRF data processing and how they differ from one another.

Dynamic Analysis. Dynamic Analysis (DA) has been developed for MA-XRF data evaluation as a response to the relatively slow iterative least squares procedures. This matrix factorization method uses an unrestricted L-LS procedure, where the abundance matrix A can be calculated as follows:

$$A = C\Gamma \tag{4}$$

Here Γ can be seen as the right inverse of the feature matrix F when comparing the DA method to the general matrix factorization formula given in Equation 3. The DA matrix $\Gamma \in^{C\times M}$ consists of M pre-defined elemental spectra. DA is therefore a supervised matrix factorization method, where the dictionary is given and the only factor that needs to be determined is the abundance matrix. However, before evaluation, it is not known which elements are exactly present. By investigating the sum spectrum, either manually or automatically, elements that are present in the painting can be examined and implemented in the DA matrix. Using the sum spectrum limits the influence of noise, however, care has to be taken since elements that are not abundantly present in the painting can be overlooked as well. Therefore, the DA matrix has to be often manually adjusted several times before the Γ and A matrices are accepted as proper decomposition matrices for the datacube C. Since the matrix Γ is physically meaningful, representing the elements, the abundance matrix A becomes a direct representation for the intensity per element per pixel; allowing us to visualise elemental distribution maps instantly. DA is the processing method adopted in Datamuncher.[9] [24]

The spectra provided by the DA matrix include contributions from bremsstrahlung, moreover, they correct for overlapping peaks by adding negative values to certain elemental profiles. A downside to the method is that it assumes fixed ratios between the fluorescence lines of each element. This complicates the detection of elements whose photons are partly absorbed.[24] Yet, this downside is negligible with the advantage DA has over iterative least squares procedures when it comes to processing speed.[24]

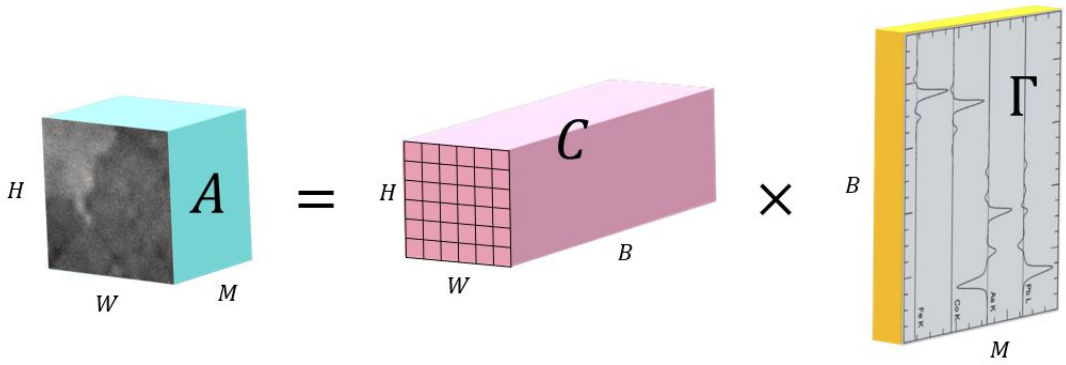


Figure 5: Schematic overview dynamic analysis. Here block A represents the abundance matrix, block C represents the hyperspectral datacube and block Γ represents the DA matrix. The inversed spectra visible on block Γ obtained from the 2015 work of Alfeld and Janssens.[24]

Unsupervised Matrix Factorization. Unsupervised matrix factorization differs from DA by not using a predefined feature matrix that represents the elements present in the painting. The feature matrix F and the abundance matrix A are formed simultaneously. Every feature in F is obtained by finding similar elemental compositions within the data. This implies that the features in F represent a combination of elements rather than a spectrum of one element. These kinds of methods highlight elemental correlations within a sample, instead of distinguishing between elements. It is important to mention that the merit of matrix factorization for pre-processing purposes is often omitted in previous research. This was pointed out by the recently published work by Chopp *et al.*, who stated that the spectral denoising capacities of these factorization methods seem to be largely unexplored.[9]

Initially, Principal Component Analysis (PCA) was often used as a factorization method for MA-XRF data. The method decomposes the data based on finding eigenvectors that cover the highest variance within the data. However, it is not always easy to interpret these vectors, which formed the feature matrix F. The components of PCA consist of negative values as well, while the datacube is entirely non-negative. Hence, the results of PCA are difficult to interpret and, moreover, do not point directly to the presence of certain elements.[26]

To solve for negative values, it was determined that both the feature matrix and the abundance matrix had to be constrained, consisting solely of positive values. The solution took the shape of non-negative matrix factorization (NMF), its implementation for MA-XRF data analyses being proposed by Alfeld *et al.*[26] Where in PCA, the datacube is decomposed in such a way that the datacube can

be reconstructed in its entirety by multiplication of the feature and abundance matrix, for NMF, like for DA, this is not the case. Due to the linearity constraints in DA and the positivity constraints in NMF, the matrix factorization of F and A will always be an approximation of the datacube C. However, where in DA only the abundance matrix could be optimised, in NMF, the least-squares approximations optimize the loss function for the feature matrix F as well as for the optimum abundance matrix A. A loss function that is frequently used for NMF is the Fast Non-Negative Least Squares (FNNLS), which solves the loss function for A and F as follows:[27]

$$A^{k+1} \longleftarrow \underset{A^k \geq 0}{\arg\min} \|C - A^k F^k\|_F^2 \tag{5}$$

$$F^{k+1} \longleftarrow \underset{F^k \geq 0}{\arg\min} \|C - A^{k+1} F^k\|_F^2 \tag{6}$$

In Equation 5, the abundance matrix A is optimized while feature matrix F is held constant, similarly, in Equation 6, the feature matrix F is optimized while abundance matrix A is held constant. NMF, like PCA, can be used to denoise the dataset.[9]NMF has, moreover, been used as a post-processing method to study Rembrandt's *Saul and David*.[26]

The process of obtaining the feature matrix F, in both PCA and NMF, is entirely data-driven. Another method called Simplex Volume Maximisation (SiVM),[28] uses a set of archetypes for the feature matrix instead. The measured datacube C is then approximated by a linear combination of actual data points within the set, representing the extremes.[4] SiVM has been applied to Hans Memling's *Portrait of a Man, possibly from the Lespinette Family*. This study is also one of the limited accounts of showing the denoising characteristics of factorization methods.[29]

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS). MCR-ALS is a version of the NMF method that allowes for various constraints to be included. Chopp *et al.* mentioned that an often-used constraint for this method was that the feature matrix F should be a combination of individual elemental spectra. They formulated the resulting loss function as follows:[9]

$$B^* A^* = \underset{B, A^k \geq 0}{\arg\min} \|C - A(MB)\|_F^2 \tag{7}$$

where the feature matrix $F = MB$ and $B \in \mathbb{R}_+^{M \times N}$ is the mixing matrix that describes the linear combination of elemental spectra present in matrix M. Hence, M is in this case actually similar to the feature matrix used during dictionary learning. As such, MCR-ALS also requires a preliminary estimate of the number of elements in the painting and their corresponding spectra. When not focused on

elemental decomposition, but rather on pure components, an initial estimate for the number of pure components present in the painting could be made by using PCA or singular value decomposition (SVD).[20] MCR-ALS has been applied to the study of Mondriaan's *Broadway Boogie Woogie*,[22] and Jackson Pollock's *Number 1A, 1948*.[20]. In the latter case, both the identification of elements as finding how those elements were combined in the paint was done using MCR-ALS.

2.1.3 Preliminary Denoising

Pre-Processing

Determining the state-of-the-art in denoising is not trivial, for not much is written on the subject. Still, in the evaluation of MA-XRF data, denoising methods are used.[9] In this section, I will touch upon these methods.

Poisson Scaling. MA-XRF data has a Poisson nature and, hence, Poisson scaling has been used to pre-process datasets. Data having a Poisson nature is characterized by noise proportional to the square root of the signal. Therefore, Poisson scaling is done so that the level of noise for each variable is equal to one another. This is done by giving each element in the dataset a weight which is consistent with its uncertainty.

Martins *et al.* mentioned that they used Poisson scaling to improve the signal-to-noise ratio when investigating Jackson Pollock's Number 1A, 1948.[20] A formulation of what happens during Poisson scaling can be written as follows:

$$\tilde{\mathbf{C}} = \mathbf{G}(\mathbf{C} - \mu\mathbf{1}\bar{\mathbf{s}}^T)\mathbf{H} \tag{8}$$

Here, $\tilde{\mathbf{C}}$ is the scaled data, $\bar{\mathbf{s}}$ is the mean spectrum and $\mathbf{1}$ is a vector of ones that is multiplied by μ which is either 0 or 1 depending whether the data is mean-centered or not. Furhter, $\mathbf{G}$ and $\mathbf{H}$ are diagonal matrices that scale the rows and columns of $\mathbf{C}$, respectively. The pixel data is scaled by making $\mathbf{G}$ equal to the inverse of the row sums of $\mathbf{C}$ and making $\mathbf{H}$ an identity matrix. During Poisson scaling, the diagonals of $\mathbf{G}$ and $\mathbf{H}$ are the inverse square roots of the mean image and mean spectrum, respectively. In other words, each column of the spectral data matrix is divided by the square root of the mean image and, further, dividing each row of the matrix by the square root of the mean spectrum.[30]

Dwell Time Variations solver. Dwell time variations were explained in section 1.2.2. Only recently this kind of noise was addressed as a distinct form of noise and tackled distinct from general noise. Dwell time variations can be corrected by normalizing for live time per pixel, which has been done by Alfeld *et al.*[23] They use the zero peak of a spectrum to determine how long a pixel is scanned.

Subsequently, by integrating over this peak, the intensity can be found, which is used as a reference for the normalization procedure.

Pre-treatment by Kogou *et al.* As mentioned, denoising steps are often used, yet rarely specified in MA-XRF. However, Kogou *et al.*[31] did elaborate on their pre-treatment of the XRF data. They applied several denoising steps before they started evaluating the data, namely a spatial median filter, using a kernel of 3 by 3, and they, furthermore, reduced the number of channels by including only the sum of five channels around each channel. These preliminary steps allowed the reduction of data, which would optimise processing efficiency,increasing the signal-to-noise ratio.

2.1.4 Background Correction

Pre-Processing or Processing

Unlike with preliminary denoising, a lot has been written about background correction. As mentioned in section 1.2.2, the background consists of bremsstrahlung and scattered primary radiation. This noise, which is clearly visible in the spectrum, can be removed in various ways. Alfeld and Janssens mentioned in 2015 that there are three ways to remove the background radiation. The first method regards the shape of the background of the sum spectrum as continuous throughout the scan and removes this shape from every spectrum. The second method finds the background through filtering procedures and removes the background on a spectrum by spectrum basis before data evaluation (processing) takes place. The third method models the background during the processing.[24] Below, the three methods are discussed in more detail.

Background Shape of the Sum Spectrum. Removing the background based on the shape of the sum spectrum is an easy and quick method to get rid of the background. However, it does require the sample to have little or homogeneous scattering. When, for example, a canvas painting is scanned having a wooden stretcher, the amount of scattering differs quite significantly between the areas with and without the stretcher behind it. In that case, this kind of background correction would not suffice.[25]

Estimating the Background on Spectrum by Spectrum Basis. One can also estimate the background and consequently subtract the background of each individual spectrum before fitting. Statistics-sensitive Non-linear Iterative Peak-clipping (SNIP) is a method that finds the background through the cutting of the peaks.[32] This is done by iteratively comparing the channel value B_i to the average

of two channel values that are n steps to the left and right of B_i. The minimum value is then substituted for the original value:

$$B_i = min[B_i, \frac{B_{(i+n)} + B_{i-n)}}{2}] \tag{9}$$

The initial value of n is often set as twice the full width at half maximum of the peaks. Every iteration this value becomes smaller and, as a result, the background can be found. This method is good to estimate irregular background shapes, however, is slow due to all the iterations that need to be performed. [24]

Background Modelling. By background modelling, we refer to the practice of background being included during the fitting of individual spectra. Typically linear or exponential polynomials are used. Linear background modelling can be described by the following formula:

$$(B_i = \sum_{d=0}^{d} a_d (i - i_0)^d) \tag{10}$$

And exponential background modelling:

$$(B_i = exp(\sum_{d=0}^{d} b_d (i - i_0)^d)) \tag{11}$$

Where B_i is the background at channel i, d stands for the degree of the polynomial, a_d and b_d are the linear and exponential parameters, respectively, and i_0 is the central channel of the fitted region. Alfeld and Janssens mention that, although modelling the background is faster and less affected by a low SNR than SNIP, the increase of parameters makes it a less robust method for background determination. Exponential polynomials are better in finding complex background shapes than linear polynomials, [33] however, polynomials with exponential variables should, according to Alfeld and Janssens solely be used for spectra with good signal-to-noise ratio. Linear variables could be used for determining the background of individual spectra, with somewhat lower statistics as well; although the statistics of these spectra still determined how high the degree of the polynomial could be. A balance has to be found between the degree being too low, not correctly rendering the background shape, or the degree being too high, resulting in the fitting model being not robust. [24] Today, removing the background is generally done during processing with the DA-method. Here, the feature matrix takes into account background contributions by using the background shape of the sum spectrum to fit each individual spectrum.

2.1.5 Absorption Correction

Post-Processing

In 1.2.2, we discussed the causes and effects of absorption. In this section, we will explain the current methods that are used to correct for artefacts that arise due to absorption.

Most MA-XRF research that focuses on removing artefacts due to absorption uses a fundamental parameter approach. In this approach, the detected photon intensity is mathematically corrected to account for factors that might influence the amount of radiation that is emitted. This entails correcting for self-absorption A_{ijk} and absorption in covering layers A'_{ijk}. The intensity R of element i, shell j and line k (for example α or β) is then formulated as:

$$R_{ijk} = m_i \cdot Y_{ijk}(E_0) \cdot A_{ijk}(E_0) \cdot A'_{ijk}(E_0) \tag{12}$$

where m_i denotes the mass of element i present in the volume, Y_{ijk} the fluorescence yield, and E_0 the energy of the incident beam. A'_{ijk} can be formulated as follows:

$$A'_{ijk}(E_0) = exp(-\chi(E_0, E_{ijk})\rho d) \tag{13}$$

with

$$\chi(E_0, E_{ijk}) = \mu(E_0)\frac{1}{sin(\alpha)} + \mu(E_{ijk})\frac{1}{sin(\beta)} \tag{14}$$

Here χ is the value for the geometrically corrected mass attenuation coefficient μ, E_{ijk} is the fluorescent energy of element i, shell j and line k, and ρd_{abs} is the areal density of the absorbing layers. In Figure 6, the geometrical dependence of the absorption is visualised: the depth of the layer and angle of incidence/take-off angle determines the amount of absorption that occurs. Note that in these simplified formulas, an assumption is made that the layers are homogeneous.[18] [25] [34] A more detailed description of the fundamental parameter approach can be found in the work by Janssens *et al.*[34]

An example of the fundamental parameter approach being used to correct for absorption-affected data is the 2013 research by Alfeld *et al.*[18] Wróbel *et al.* followed up on this research in 2016 by adapting these methods to allow absorption correction for polychromatic excitation. Next to that, they adapted the process of parameter selection by Alfeld *et al.* in order to reduce human input.[35] Recently a study by Alfeld *et al.* focussed on another absorption-affiliated problem, where they used the fundamental parameter approach in order to correct for varying distances between the measurement head and the sample. They combined the fundamental parameter approach with information of the Ar signal, which is proportional to the working distance.

Yet, absorption remains a difficult aspect to correct for. Since, as was the case in the above-stated examples, fitting of the data is done first.

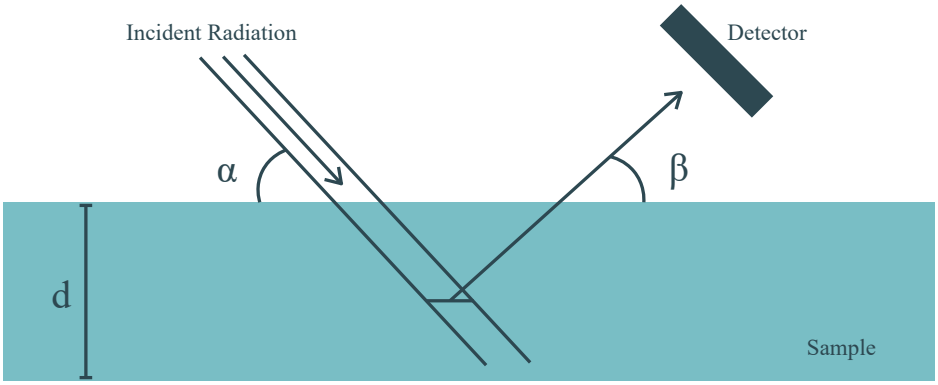


Figure 6: Detection Geometry used for Calculating the Absorption Corrected Photon Intensity. Here α is the angle of incidence for the primary radiation, β is the take-off angle of the emitted radiation and d the sample thickness.

2.2 Research Gap

In the previous chapters, I described the working principle of MA-XRF and the state of the art in data processing. What is striking is that little attention is given to the spatial relations within the dataset. Recently, however, more research on the use of spatiality in MA-XRF is being published. In this chapter, a quick overview of these methods will be given, after which an evaluation of these methods is made; herewith introducing the research gap that will be the basis for my research.

2.2.1 Current Research on the use of Spatiality in MA-XRF Processing

In 2022, Chopp *et al.* published an article that reviews image processing methods that could be applied to MA-XRF data to improve data evaluation. These image processing methods differ from current MA-XRF processing procedures by taking spatial relations between pixels into account. Chopp *et al.* ascribed the limited use of spatial information when fitting or denoising MA-XRF data to the fact that MA-XRF data has been seen as a collection of spectra rather than a collection of images.[9] Therefore, Chopp *et al.* aimed to bring knowledge on methods that use spatiality to a wider audience in order to instigate further research into MA-XRF image processing. In this section, a short overview of the discussed methods is given, complemented by two additional sections on spatial processing methods used in MA-XRF. One that discusses neural networks and the other discussing a second work by Chopp *et al.*

PCA with k-means Clustering. Vekemans *et al.* developed one of the first data evaluation methods that used spatial as well as spectral information on MA-XRF data by combining PCA with k-means clustering.[36] PCA was used to generate eigenimages describing the data and k-means clustering was used afterwards to find pixels with similarities in order to find distinct regions with similar elemental composition. By generating the sum spectrum of these distinct regions, trace elements were more likely to be detected than when the sum spectrum of a sample in its entirety was used.

Neural Network Clustering. In more recent years, neural networks are being used for data evaluation. Kogou *et al.* developed an unsupervised machine learning algorithm, based on the Self Organising Map (SOM) method, which uses spectral as well as spatial information in order to find clusters consisting of similar spectra.[31] After the clusters have been formed, just as the PCA method by Vekemans *et al.*, the sum spectra of those clusters could be used to improve the identification of elements, particularly trace elements.

Neural network clustering is a method where in consecutive lines of neurons, the redundant input is transformed into a sparse cluster representation. For every consecutive line, the number of neurons is less than the number in the previous line. The input layer of neurons is formed by the pixels and their respective spectra, while the output consists of the clusters. The algorithm is trained from the dataset at hand, hence, no reference is necessary. The only input that was required for the algorithm by Kogou *et al.* was the number of clusters.

The advantage of this method over current data-evaluation methods, as described in section 2, is the fact that a larger number of elements and/or complex material compositions can be evaluated as well. Moreover, trace elements can be more easily detected.

Richardson-Lucy Deblurring. The Richardson-Lucy method, a super-resolution method based on the Fourier transform, super-resolves the elemental distribution map Z by convolving the acquired elemental map L with a known point spread function (PSF) P. This method can iteratively deblur distribution images by converging at the maximum likelihood solution. The updated distribution map can be determined as follows:

$$Z^{k+1} = Z^k \cdot (\hat{P} * \frac{L}{P * Z^k}) \tag{15}$$

Here, $\hat{P}$ is the flipped matrix P which is obtained by reversing the order of the elements in the rows and columns of P. Contradictory, the Richardson-Lucy method, being notably used to deblur shot noise-polluted images, is according to the authors only useful for deblurring elemental distribution maps of elements with high photon counts.[37]

Super-Resolution. Super-resolution uses spatial coherency between pixels to improve the spatial resolution of the data. This method allows fewer pixels to be scanned by estimating sub-pixels.

Dai *et al.* used a high-resolution RGB image $R \in [0,1]^{3\times H_s \times W_s}$ to improve the transformation from low-resolution data $X \in^{C\times H_l, W_l}$ into super-resolved data $Y \in^{C\times H_s \times W_s}$.[38] They separated the MA-XRF data in visual v and non-visual nv components so that the visual component could be directly linked to the RGB image. The matrix decomposition they applied was as follows:

$$Y_v = F_v^{xrf} A_v \tag{16}$$

$$Y_{nv} = F_{nv}^{xrf} A_{nv} \tag{17}$$

$$R = F^{rgb} A_v \tag{18}$$

Where F is the feature matrix and A the abundance matrix. Furthermore, the abundance matrix A_v is the same for the RGB decomposition as for the visual component of MA-XRF. Low-resolution data X and super-resolved data Y are related to one another by:

$$X_v = Y_v \cdot S \qquad X_{nv} = Y_{nv} \cdot S \tag{19}$$

where $S \in \{0,1\}^{1\times H_s \times W_s}$ is a binary sampling matrix. The optimization problem then becomes:

$$\underset{\Theta\geq 0,\quad \|A\|_0 \leq s}{\arg\min} \|X-(F_v^{xrf}A_v)\cdot S-(F_{nv}^{xrf}A_{nv})\cdot S\|_F^2+\|R-D^{rgb}A_v\|_F^2+\lambda_{TV}TV(D_{nv}^{xrf}A_{nv}) \tag{20}$$

with $\Theta = \{D_v^{xrf}, D_{nv}^{xrf}.D^{rgb}, A_v, A_{nv}\}, A = A_v + A_{nv}$. It can be seen that a total variation (TV) regulizer is used which exploit neighbouring pixel correlations.

Inpainting. Inpainting is similar to super-resolution. However, it differs from super-resolution because the scanning grid for inpainting is not uniform.

Dai.*et al.* adapted their super-resolution method to be applicable in inpainting procedures as well, for inpainting allows the creation of an optimal sampling strategy.[39] Their adaptation to inpainting was achieved by adding an optimization constraint saying that the gradient of the RGB image should be the same as the gradient for the MA-XRF image. A further improvement was achieved by Yan *et al.* who separated the data in common and uncommon features rather than visible and not-visible, herein also the RGB data was decomposed.[14]

Multi-Modal Dictionary Learning with a Poisson Noise Model. Where one paper by Chopp *et al.* generated an overview of currently existing image processing techniques that could be applied to evaluate MA-XRF data, another paper by the same research group applied some of these methods. Their work aims to increase the signal-to-noise ratio of scans that were obtained by using a low dwell time by applying a dictionary learning method with a Poisson noise model together with comparing the MA-XRF data to a RGB-image.[7]

In the process of fitting a dictionary, usually, a least squares approach is used as an objective function, as became apparent from section 2. Chopp *et al.*, however, apply the Poisson negative log likelihood (PNLL) as a loss function. This method is particularly convenient for data with a low photon count since the least squares approach assumes a Gaussian distribution of the data. The PNLL can be described as follows:

$$\mathcal{P}(F,A) = \sum_{c=1}^{C}\sum_{n=1}^{N}(FA)_{c,n} - \mathbf{C}_{c,n}\cdot ln(FA)_{c,n} \tag{21}$$

where $\mathcal{P}$ is the data fidelity term, $\mathbf{C}$ is the obtained datacube, F the (predetermined) feature matrix and A the abundance matrix. The PNLL being the objective function, minimizes the Poisson negative log likelihood estimation.

Chopp *et al.*, furthermore, used a TV regularizer to smooth the data. Their TV regularizer was variable with the RGB image; making use of the fact that local areas having similar colours, most likely also have similar elemental profiles, and vice versa. With the addition of a sparsity constraint ℓ_0, the combined denoising and fitting procedure could be written as follows:

$$F^*, A^* = \underset{F,A\geq 0}{\arg\min}\, \mathcal{P}(F, A) + \lambda_{TV} TV(A) + \lambda_{l_0} \|A\|_0 \tag{22}$$

Where λ_{TV} is a positive scalar.

This method uses the information of neighbouring pixels by implementing the TV regularizer which eventually resulted in outperforming the MCR-ALS method in denoising. Moreover, better detection of elements that are little present in the painting was possible.[9]

In the consequently published review paper by Chopp *et al.*, the TV regularizer as applied in Equation 22 was incorporated in the future outlook, being proposed as a spatial denoising method that could not merely combine RGB images with neighbouring pixel information, but any kind of prior information.[9]

2.3 Research Proposal

This literature review started by stating the problem of MA-XRF requiring long scanning times. The hypothesis was that the long scanning time, being a result of the long dwell time per pixel, could be resolved by increasing the signal-to-noise ratio. After the state-of-the-art data evaluation methods were discussed in section 2, it became clear that few researchers have discussed pre-processing methods for MA-XRF. The only widely applied method to increase the SNR is Poisson scaling, meanwhile, all the other processing steps are mainly focused on fitting the data. Chopp *et al.*,[9] commented in their recent work that taking into account the Poisson nature of the data and upping the SNR should become a higher priority, pointing out that there are two main ways to limit the effects of noise: either increasing the dwell time, or apply pre-processing methods. Following this statement, Chopp *et al.* opted for using spatiality in order to improve data-evaluation. Hence, by applying pre-processing methods, furthermore using local spatial relations, the dwell time, and herewith the overall scanning time, could be reduced.

In the previous section, several methods which exploit spatial relations to improve the evaluation of XRF spectra were discussed. Here, the clustering methods described in 2.2.1 were used to improve the data-evaluation by regarding sum

spectra of clusters instead of the entire painting. These clusters, having more similarities, allowed for more precise detection of the elements and detection of trace elements. Although the denoising strategies were quite successful, PCA remains a method that creates an output that has no physical significance and, therefore, is hard to interpret.[26] Furthermore, neural network denoising is not ideal for MA-XRF data, since there is limited training data.[9]

Except for Multi-modal dictionary learning with a Poisson noise model, other methods discussed in section 2.2.1 do not focus on improving the SNR. The Richardson-Lucy method seeks to compensate for a larger beam size, which might improve the spatial resolution of elemental distribution images but does not do so much to improve the spectral resolution in the process. Super-resolution and inpainting allow for longer dwell times since fewer pixels have to be scanned, hence, spectra with better statistics can be obtained. However, a downside to these methods is that it is mostly based on RGB imaging. Therefore, merely improvements in spectral decomposition are made for the visible layers, while the MA-XRF signal has a significant sub-surface contribution.

It is essential for the goal of reducing the scanning time that the Poisson nature of the data is taken into account; for shot noise is the most prominent kind of noise in data that is scanned with low dwell times. Multi-model dictionary learning with a Poisson noise model is one of the first denoising strategies taking into account shot noise and applies spatial denoising strategies to improve the spectral SNR. A downside, however, is that also this procedure uses RGB images.

Hence, although spatial relations could be used to improve the data-evaluation and even account for shot noise, not all spatial denoising methods increase the *spectral* SNR and, the few that do, depend on prior knowledge such as RGB images or require training data.

Like the work by Chopp *et al.*,[7] our research focuses on removing shot noise from MA-XRF data, using the information of neighbouring pixels. Moreover, we use the information of neighbouring channels. The necessity of RGB images is circumvented by creating a filter that needs no prior information. To this extent, wavelet filter bank denoising is proposed as a suitable method. This research will fill the gap in the application of denoising methods in MA-XRF, moreover, it will use the, in MA-XRF relatively unexplored but vast field of image processing for increasing the SNR.

3 Wavelet Filter bank Denoising

The research objective of reducing the necessary dwell time to attain a specific SNR has been translated into the goal of finding a denoising method that reduces the noise floor by exploiting local spatial relations within the MA-XRF data. There are a lot of different denoising methods that spatially denoise data, however, not all of these methods are suited for MA-XRF data denoising. In order to find a proper denoising method, MA-XRF-specific data characteristics need to be taken into account. In this chapter, those data characteristics are explained and wavelet filter bank denoising is proposed and explained as a suitable method to denoise MA-XRF data. First, the choice of wavelets as a suitable tool for denoising is substantiated. Hereafter, the mathematical background of wavelet filter bank denoising is described. Finally, the design for a wavelet filter suitable for MA-XRF denoising is explained.

3.1 Why Wavelet Denoising?

In this section the signal properties of MA-XRF data are described, after which the choice for wavelet denoising is explained.

3.1.1 Signal Properties

Heterogeneous Signal Structure. The objects scanned with MA-XRF have a wide variety of material compositions. Also, between similar objects large differences in material composition can be found. You can even find, for example, painters that within their own oeuvre change materials as they and others experiment.[40] Therefore, a denoising method suitable to denoise the wide variety of MA-XRF data cannot lean on too much prior information on the expected content of the signal and thus, the denoising should be as data-driven as possible.

Localized Signal. The spatial data from objects scanned by a MA-XRF scanner has a localized nature, where elemental distributions are most of the time local to a specific area and discontinuous at the edge with neighbouring areas. These local distributions tend to be smooth internally and their edges are described by abrupt changes in photon intensity. Therefore, a denoising method that denoises MA-XRF data should avoid removing these high-frequency edge-describing components. In this respect, denoising methods that are applied in denoising visual light photographs, as those mentioned by Chopp *et al.*,[9] could be used.

Spatial and Spectral Relationship. Although methods that are commonly used for image denoising would do reasonably well in denoising the elemental distribution images, this is not the objective of this research. Here, the reason for exploiting spatial relations within the hyperspectral datacube is to improve the *spectral* SNR (i.e., the SNR along the spectral axis). The difference between denoising a single gray-level image and spatially denoising a MA-XRF dataset is that the absolute values of the intensity of the pixels have a physical meaning. It is not relevant to simply smooth out the dataset on an image-by-image basis since then the spectral signal, covering many channel images, would change significantly. It is, therefore, necessary to find a *spatial* denoising method that nevertheless does not distort the *spectral* signal in the process.

Photon Shot Noise. The aim of this research is to improve the SNR by lowering the noise in order to lower the necessary dwell time by requiring less signal for the same SNR. For this purpose, it is important to know what kind of noise is most prominent in the case of low dwell times. In section 1.2.2, various kinds of noise that are present in a MA-XRF scan were discussed. Noise, such as the background radiation, escape peaks, and pile-up peaks will always be there, no matter if you apply a long or short dwell time. These kinds of noise are, furthermore, already well accounted for in standard MA-XRF processing procedures. However, the noise that constitutes the main limitation when the dwell time is shorter, the photon shot noise, is often not taken into account. As explained in 1.2.2, photon shot noise is a result of the discontinuous flow of photons, which, when you scan for a short time, might result in strongly fluctuating measurements. These fluctuations make it difficult to interpret the data and, therefore, it is necessary that the denoising algorithm reduces the effect of shot noise.

3.1.2 Choosing for Wavelet Denoising

Based on the data characteristics above, a customized wavelet-based denoising algorithm was proposed, more specifically based on perfect reconstruction wavelet filter banks.[41]

Wavelet filtering, unlike the currently popular neural networks, does not require *a priori* training information. The data itself can be used to find proper thresholding values for denoising.[42] Moreover, wavelets have been commonly applied in image denoising. Wavelets have both a high spatial and high frequency resolution, through which they are well suited to detect local signal features within an image, such as abrupt intensity changes. Furthermore, wavelets are generally well-suited to filter out noise while keeping the important edge-describing high-frequency components of images intact.[43]

As previously mentioned, in MA-XRF data every spatial intensity value is also a spectral one, and we cannot smooth the data in space without impacting the spectral dimension as well. Wavelet decomposition does not smooth the data but instead redistributes the data, hence, the number of photons within every channel remains constant; this prevents the distortion of the spectral signal. In the general wavelet filterbank decomposition, the spatial signal is approximated using a decomposition into approximation coefficients and detail coefficients. The approximation coefficients represent the average signal of neighbouring pixels and, if only these would be stored, some information from the spatial signal would be lost. The detail coefficients contain the values of the deviation from the signal approximation encoded by approximation coefficients. By summing the approximation signal and detail signal, encoded by the approximation and detail coefficients respectively, the signal can be completely reconstructed. In other words, using wavelet filter bank decomposition, a measured signal can be decomposed into an approximation part and a details part, while maintaining the ability to reconstruct the original measurement. This makes it quite suitable as a spatial denoising method for MA-XRF data. However, please note in its traditional form wavelet filterbank decomposition is only aware of the spatial domain and the spectral dimension is largely ignored. Awareness of the latter requires building spectral awareness into the thresholding function, a customization developed in this thesis.

After the signal is decomposed, it is possible to threshold the detail coefficients since eliminating detailed variation in the measurement can amount to removing small detail noise. A lot of different threshold determination methods have been developed. Most of them assume Gaussian noise in the measurement. There are some threshold determination methods that do account for shot noise, however, they also apply filters before the signal is decomposed into approximation and detail coefficients. This tends to nullify the advantageous properties of wavelets in the preservation of all the important MA-XRF signal properties during decomposition. Since existing thresholding methods did not provide the necessary aspects for MA-XRF denoising, e.g. in terms of spectral awareness, a novel procedure needed to be designed to remove the Poisson noise by thresholding wavelet coefficients in a spectrum-aware manner.

3.2 Wavelet Filter Bank Denoising, Theory

Wavelet filter bank denoising is done in three consecutive steps:

1. The signal is decomposed in a pre-specified number of levels, using the discrete wavelet transform. For each level, detail and approximation coefficients are obtained through decomposing the signal (which is the measurement at level 1 and the approximation signal in subsequent levels) with a high- and

low-pass filter respectively.

2. The detail coefficients of every level, are thresholded to remove the small detail noise variation. The thresholding procedure entails retaining detail coefficients above the threshold as-is, while detail coefficients that fall below are replaced by zeros.

3. The signal is reconstructed, using the inverse discrete wavelet transform. The thresholded denoised detail coefficients and the approximation coefficient are transformed back and summed in order to obtain the denoised version of the original measurement.

In this section, an overview of the theory behind wavelet filter bank denoising is provided.

3.2.1 Signal Analysis & Synthesis

In the first step of wavelet filter bank denoising, the signal of interest is decomposed. The decomposition procedure is similar to that of the Fourier transform, where a signal is decomposed into its complex exponential components.Yet, where the Fourier transform decomposes the signal in (continuous) functions on an infinite support, the wavelet transform can decompose a signal into (discontinuous) signals on a finite support. The finite support basis of the decomposition allows the accurate localization of signals not only in the frequency domain, but also in the time/space domain.

The Wavelet. As previously mentioned, wavelets are able to accurately analyse localized signals on a finite support. Moreover, since there are various different shapes of wavelets, a broad range of signals can be analysed and approximated efficiently. Examples of wavelets are shown in Figure 7.

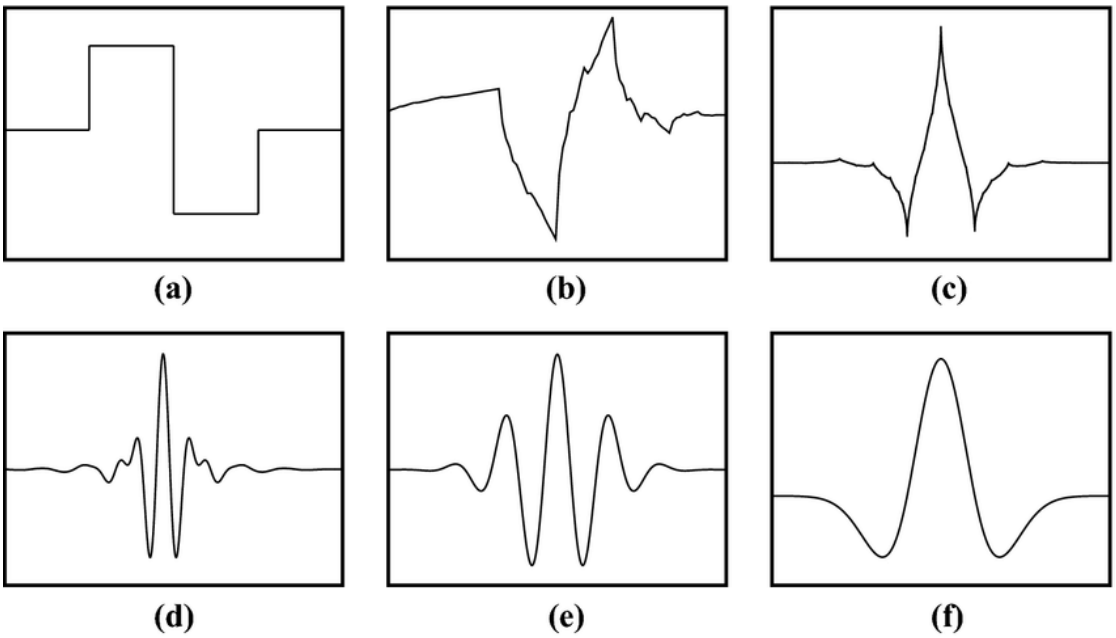


Figure 7: Examples of wavelets (a) Haar (b) Daubechies-2 (c) Coiflet-1 (d) Meyer (e) Morlet (f) Mexican Hat.[44]

Wavelets are signals on a finite support that can be scaled and translated. Through scaling, high and low frequencies can be captured and through translation, the correlation of the signal with the wavelet at any time/space instance can be determined. Every wavelet represented in Figure 7 can be seen as a mother wavelet $\psi(x)$ from which the scaled and translated functions are derived as follows:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi(\frac{x-b}{a}) \tag{23}$$

Here, a is the scaling coefficient and b is the translation coefficient.

The Wavelet Transform. The set of coefficients that are obtained by taking the inner product of the signal $f(x)$ with wavelet $\psi_{a,b}$ for a varying a and b is called the wavelet transform. By transforming the signal $f(x)$ to another domain, it becomes easier to analyse the signal. The wavelet transform can be written as follows:

$$C(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(x)\Psi(\frac{x-b}{a})dx \tag{24}$$

Taking the inner product of the signal with daughter wavelets with different scales and translations, varying with nearly infinitesimal stepsize, allows the signal to be represented with very high accuracy. However, doing this for a very small step-size is often computationally heavy due to an enormous number of coefficients being produced. Therefore, in practice, often a discrete version of the wavelet transform is used, where scaling and translating are done at specific intervals.[45] The discrete wavelet transform can be written as follows:

$$D[a,b] = \frac{1}{\sqrt{a}} \sum_{n \in \mathbb{Z}} f[x_n] \psi[\frac{x_n - b}{a}] \tag{25}$$

The interval used is often dyadic; here the scaling and translation coefficients are scaled with a power of two:

$$a = 2^j \quad \text{for} \quad j = 1, 2, 3...m \tag{26}$$

$$b = 2^j k \quad \text{for} \quad j = 1, 2, 3...m \quad \text{and} \quad k = 1, 2, 3, ...h \tag{27}$$

Equation 23 can thus be rewritten as:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \psi(\frac{x - 2^j k}{2^j}) = \frac{1}{\sqrt{2^j}} \psi(2^{-j} x - k). \tag{28}$$

Here, j is the level, 2^j is the scale and $2^j k$ is the shift along the time/space axis. By scaling the translation coefficient and the scaling coefficient by a factor of two, it is possible to create an orthonormal wavelet basis.[46] Orthonormality allows for a sparse representation of the signal, where the information that is stored in each wavelet coefficient is unique.

Wavelet Filter bank. In wavelet filter bank analysis, an orthonormal wavelet basis is used to decompose a signal into high-frequency and low-frequency components, whereafter both high-frequency and low-frequency components are downsampled by two-fold. Then, the low-frequency component can be filtered once more into a low- and high-frequency component. This can be done for several consecutive *levels*. Each time, the resulting low- and high-frequency components are downsized by a factor two, in doing so, applying dyadic scaling. This is visualised in Figure 8.

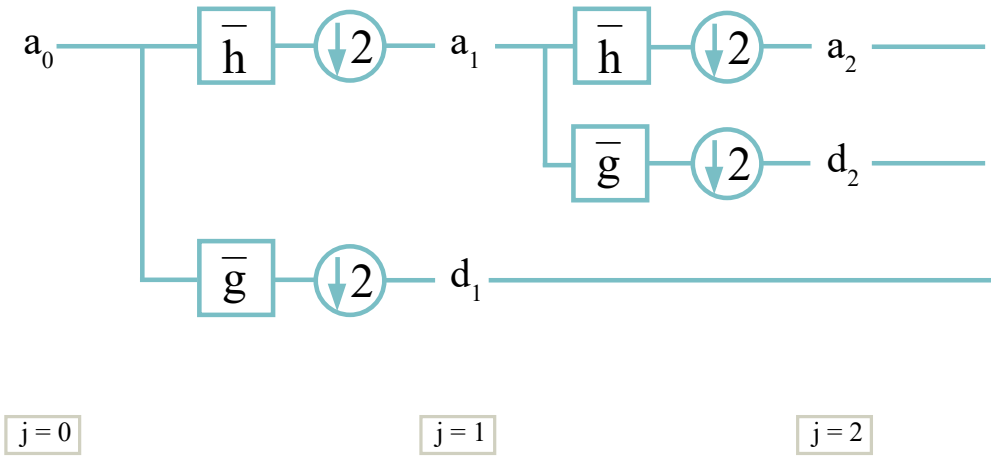


Figure 8: Wavelet Filter bank Analysis for a 2 Level Decomposition. Here, a_0 denotes the input signal, $\bar{h}$ and $\bar{g}$ are the decomposition low-pass and high-pass filters, respectively. a_1 and a_2 denote the approximation coefficients of the first and second decomposition level, respectively. Similarly, d_1 and d_2 denote the detail coefficients of the first and second decomposition level. j indicates the level.

The working principle of wavelet filter banks is derived from the multiresolution theory of Mallat and Meyer. In order to apply multiresolution analysis, a new function has to be introduced, the *scaling function* $\phi_{j,k}$. Similarly to the wavelet functions in Equation 28, scaling functions can be defined as a family of functions that are generated from a father function by scaling and translating:[47]

$$\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\phi(2^{-j}x - k) \tag{29}$$

Then, the low-frequency components of each level j are represented by approximation coefficients $a_{j,k}$, which are obtained by taking the inner product of the original signal a_0 with the *scaling function* $\phi_{j,k}$. The high-frequency components are represented by detail coefficients $d_{j,k}$, which are obtained by taking the inner product of the original signal a_0 with the *wavelet function* $\psi_{j,k}$. Here, the scaling function functions as a low-pass filter and the wavelet function as a band-pass filter:[48]

$$a_{j,k} = \langle a_0, \phi_{j,k} \rangle; \qquad d_{j,k} = \langle a_0, \psi_{j,k} \rangle \tag{30}$$

A property of the multiresolution approximation is that each coarser resolution level space is contained within the higher resolution spaces, $V_{j+1} \subset V_j$. Where $V_j = \overline{\text{Span}_k[\phi_{j,k}(x)]}$. This entails that the scaling and wavelet functions of the lower resolution level can be derived from the scaling function at a higher resolution level, which is called the two scale-relation:

$$\phi_j = \sum_{n\in\mathbb{Z}} h_n \, \phi_{j-1,n} \tag{31}$$

$$\psi_j = \sum_{n\in\mathbb{Z}} g_n \, \phi_{j-1,n} \tag{32}$$

Here, n is the integer index, h_n and g_n are the scaling/low-pass filter and wavelet/high-pass filter, respectively. One can rewrite Equation 31 and Equation 32 to incorporate the dyadic scaling and translation coefficients:[49]

$$\phi(x) = \sum_{n\in\mathbb{Z}} h(n)\sqrt{2}\phi(2x-n) \tag{33}$$

$$\psi(x) = \sum_{n\in\mathbb{Z}} g(n)\sqrt{2}\phi(2x-n) \tag{34}$$

The derivation of the lower-resolution wavelet and scaling functions can be visualised easily using the Haar wavelet. The Haar low-pass filter h and high-pass filter g can be denoted as:

$$h = \frac{1}{\sqrt{2}}[1 \qquad 1]; \qquad g = \frac{1}{\sqrt{2}}[1 \qquad -1]. \tag{35}$$

By substituting the Equations from 35 in Equation 33 and Equation 34, it can be seen how the scaling function $\phi(x)$ and the wavelet function $\psi(x)$ of a lower resolution level can be reconstructed by the scaling functions of a higher resolution level:

$$\phi(x) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_n\phi(2x-n) = \sqrt{2}(\frac{1}{\sqrt{2}}\phi(2x) + \frac{1}{\sqrt{2}}\phi(2x-1)) = \phi(2x) + \phi(2x-1) \tag{36}$$

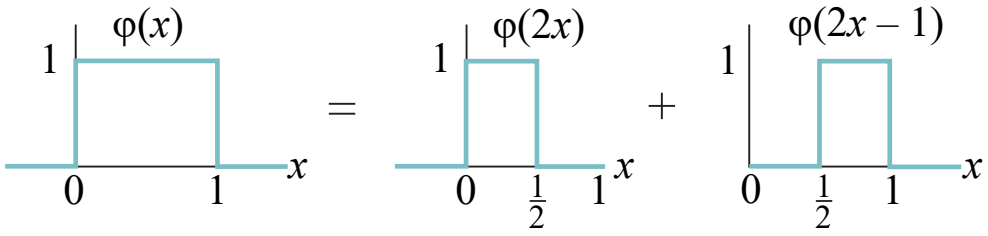


Figure 9: Constructing the Haar scaling function $\phi(x)$ of a lower resolution from the Haar scaling function of a higher resolution (dyadically scaled) according to the two scale-relation.

$$\psi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} g_k \phi(2x - n) = \sqrt{2}(\frac{1}{\sqrt{2}}\phi(2x) - \frac{1}{\sqrt{2}}\phi(2x - 1)) = \phi(2x) - \phi(2x - 1) \tag{37}$$

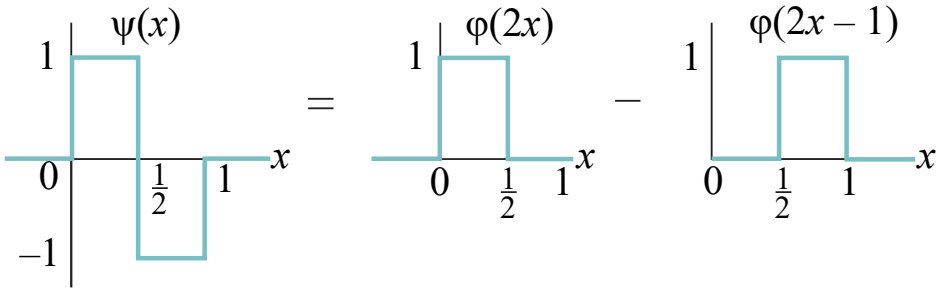


Figure 10: Constructing the Haar wavelet function $\psi(x)$ of a lower resolution from the Haar scaling function of a higher resolution (dyadically scaled) according to the two scale-relation.

In order to find the approximation and detail coefficients of each consecutive level, it is possible to rewrite Equations 30 in terms of the scaling and wavelet filters h and g.

Recall Equation 34:

$$\psi(x) = \sum_{n\in\mathbb{Z}} g_n \sqrt{2}\phi(2x - n)$$

Scaling and translating the time variable gives us:

$$\begin{aligned} \psi(2^{-j}x - k) &= \sum_{n\in\mathbb{Z}} g_n \sqrt{2}\phi(2(2^{-j}x - k) - n) \\ &= \sum_{n\in\mathbb{Z}} g_n \sqrt{2}\phi(2^{-j+1}x - 2k - n) \end{aligned} \tag{38}$$

Recall Equation 28:

$$\psi_{j,k} = 2^{-j/2}\psi(2^{-j}x - k)$$

We can rewrite Equation 28 in terms of Equation 38:

$$\psi_{j,k} = 2^{-j/2} \sum_{n\in\mathbb{Z}} g_n \sqrt{2}\phi(2^{-j+1}x - 2k - n)$$

$$\begin{aligned} \text{Take} \qquad m &= 2k + n, \\ \psi_{j,k} &= 2^{-j/2} \sum_{m\in\mathbb{Z}} g_{m-2k} \sqrt{2}\phi(2^{-j+1}x - m) \\ &= \sum_{m\in\mathbb{Z}} g_{m-2k} 2^{(-j+1)/2}\phi(2^{-j+1}x - m) \end{aligned} \tag{39}$$

Notice that Equation 28 is similar for the father wavelet ϕ and, for $j - 1$ can be denoted as:

$$\begin{aligned} \phi_{j-1,m} &= 2^{-(j-1)/2}\phi(2^{-(j-1)}x - m) \\ &= 2^{(-j+1)/2}\phi(2^{-j+1}x - m) \end{aligned} \tag{40}$$

Then, substituting Equation 40 in Equation 39 results in:

$$\psi_{j,k} = \sum_{m\in\mathbb{Z}} g_{m-2k}\phi_{j-1,m} \tag{41}$$

Recall the formula for the detail coefficients from Equation 30:

$$\begin{aligned} d_{j,k} &= \langle a_0, \psi_{j,k} \rangle \\ &= \int a_0(x)\psi_{j,k}(x)dx \end{aligned}$$

Substituting Equation 41 in the formula for the detail coefficients from Equation 30

we obtain:

$$d_{j,k} = \sum_{m\in\mathbb{Z}} \overline{g_{m-2k}} \int a_0 \phi_{j-1,m} \tag{42}$$

It can be recognised that the integral is the scaling function from Equation 30 at level $j-1$, hence, the formula from which the detail coefficients of the following level can be determined becomes:

$$d_{j,k} = \sum_{m\in\mathbb{Z}} \overline{g_{m-2k}} a_{j-1,m} \tag{43}$$

Similarly, for the approximation coefficients:

$$a_{j,k} = \sum_{m\in\mathbb{Z}} \overline{h_{m-2k}} a_{j-1,m} \tag{44}$$

This derivation has been obtained by regarding *Ten Lectures of Wavelets* by Ingrid Daubechies and *Introduction to Wavelets and Wavelet Transforms, A Primer* by Burrus *et al.*[49] [47] Note that the notation of Daubechies is used in which higher values for j indicate an increasingly *coarser* resolution.

The main advantage of wavelet filter bank denoising is the possibility of perfect reconstruction of the original signal. Perfect reconstruction is obtained when no aliasing and no distortion happens throughout the filtering of the signal.[50] This is achieved by using conjugate quadrature filters (CQF), $h[n]$ and $g[n]$, where:

$$g[n] = (-1)^n h[n] \tag{45}$$

Here, $g[n]$ and $h[n]$ are the *reconstruction* high-pass and low-pass filters, respectively. The *decomposition* filters relate to them as:

$$\bar{h}[n] = h[-n]; \qquad \bar{g}[n] = g[-n] \tag{46}$$

Through combining the theory of multi-resolution with conjugate quadrature filters, a perfect reconstruction wavelet filter bank, as shown in Figure 11, is obtained.

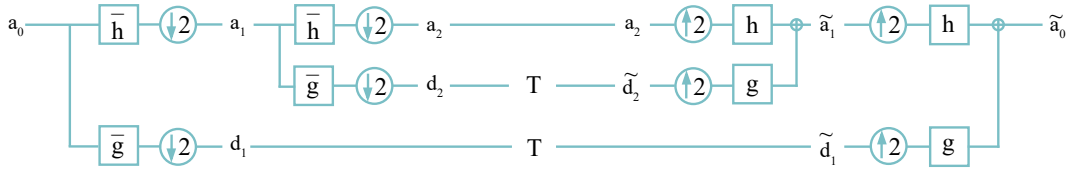


Figure 11: Analysis and Synthesis Filter bank for 2 Decomposition Levels. Here, a_0 denotes the input signal, $\bar{h}$ and $\bar{g}$ are the decomposition low-pass and high-pass filters, respectively. a_1 and a_2 denote the approximation coefficients of the first and second decomposition level, respectively. Similarly, d_1 and d_2 denote the detail coefficients of the first and second decomposition level. The T stands for thresholding. Furthermore, $\tilde{d}_1$ and $\tilde{d}_2$ are the detail coefficients after thresholding. h and g are the reconstruction low-pass and high-pass filters, respectively. $\tilde{a}_1$ is the reconstructed approximation coefficient of the first level and $\tilde{a}_0$ is the output signal.

Example of a Perfect Reconstruction Haar Wavelet Filter bank

To illustrate the theory above, here, an example is given of the decomposition and reconstruction of a signal a_0 using a perfect reconstruction filter bank with the Haar wavelet.

The vector representation of the signal a_0 is:

$$\begin{bmatrix} 4 & 2 & 3 & 1 & 4 & 2 & 3 & 2 & 9 & 8 & 9 & 7 & 9 & 7 & 8 & 8 \end{bmatrix}$$

The first stage of signal decomposition in the filter bank is as follows:

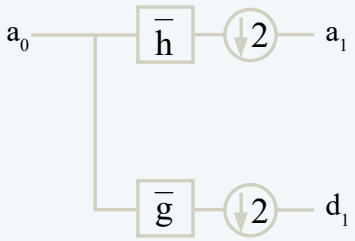


Figure 12: First Level Decomposition Filter bank.

Hence, to determine the approximation coefficients a_1 and detail coefficients d_1 we need the high-pass and low-pass *decomposition* filters of the Haar wavelet, $\bar{h}$ and $\bar{g}$ respectively. Recall Equation 35:

$$h = \frac{1}{\sqrt{2}}[1 \qquad 1]; \qquad g = \frac{1}{\sqrt{2}}[1 \qquad -1],$$

which are the Haar *reconstruction* filters. By substituting these into Equation 46, the haar decomposition filters can be derived:

$$\bar{h} = \frac{1}{\sqrt{2}}[1 \qquad 1]; \qquad \bar{g} = \frac{1}{\sqrt{2}}[-1 \qquad 1]$$

Now recall Equation 44 and Equation 43:

$$a_{j,k} = \sum_{m \in \mathbb{Z}} \overline{h_{m-2k}} a_{j-1,m}, \qquad d_{j,k} = \sum_{m \in \mathbb{Z}} \overline{g_{m-2k}} a_{j-1,m}$$

Continuation: Example of a Perfect Reconstruction Haar Wavelet Filter bank

These equations are equivalent to downsizing by a factor two the convolution of the finite impulse response filters with the approximation coefficients of a higher resolution level:[50]

$$a_{j,k} = (\downarrow 2)[a_{j-1,m} * \overline{h_{-m}}] \tag{47}$$

$$d_{j,k} = (\downarrow 2)[a_{j-1,m} * \overline{g_{-m}}] \tag{48}$$

Note that $\bar{h}$ and $\bar{g}$ are functions of $-m$ so the signal is actually convoluted with the inverse of the decomposition filters:

$$\bar{h}_m^{inv} = \overline{h_{-m}} = \frac{1}{\sqrt{2}}[1 \qquad 1]; \tag{49}$$

$$\bar{g}_m^{inv} = \overline{g_{-m}} = \frac{1}{\sqrt{2}}[1 \qquad -1]; \tag{50}$$

Hence, the approximation coefficients of the first level, a_1, are calculated as follows:

$$a_1 = (\downarrow 2)[\begin{bmatrix}4 & 2 & 3 & 1 & 4 & 2 & 3 & 2 & 9 & 8 & 9 & 7 & 9 & 7 & 8 & 8\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}]$$

$$a_1 = (\downarrow 2)[\frac{1}{\sqrt{2}}\begin{bmatrix}4 & 6 & 5 & 4 & 5 & 6 & 5 & 5 & 11 & 17 & 17 & 16 & 16 & 16 & 15 & 16 & 8\end{bmatrix}]$$

$$a_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}6 & 4 & 6 & 5 & 17 & 16 & 16 & 16\end{bmatrix}$$

And the detail coefficients of the first level, d_1, can be written as:

$$d_1 = (\downarrow 2)[\begin{bmatrix}4 & 2 & 3 & 1 & 4 & 2 & 3 & 2 & 9 & 8 & 9 & 7 & 9 & 7 & 8 & 8\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\end{bmatrix}]$$

$$d_1 = (\downarrow 2)[\frac{1}{\sqrt{2}}\begin{bmatrix}-4 & 2 & -1 & 2 & -3 & 2 & -1 & 1 & -7 & 1 & -1 & 2 & -2 & 2 & -1 & 0 & 8\end{bmatrix}]$$

$$d_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}2 & 2 & 2 & 1 & 1 & 2 & 2 & 0\end{bmatrix}$$

Continuation: Example of a Perfect Reconstruction Haar Wavelet Filter bank

This can be done for a_2 en d_2 as well:

$$a_{2,k} = a_{1,m} * \bar{h}_m^{inv}, \qquad d_{2,k} = a_{1,m} * \bar{g}_m^{inv}$$

$$a_2 = (\downarrow 2)[\frac{1}{\sqrt{2}}\begin{bmatrix}6 & 4 & 6 & 5 & 17 & 16 & 16 & 16\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}]$$

$$a_2 = \frac{1}{\sqrt{4}}\begin{bmatrix}10 & 11 & 33 & 32\end{bmatrix}$$

$$d_2 = (\downarrow 2)[\frac{1}{\sqrt{2}}\begin{bmatrix}6 & 4 & 6 & 5 & 17 & 16 & 16 & 16\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\end{bmatrix}]$$

$$d_2 = \frac{1}{\sqrt{4}}\begin{bmatrix}2 & 1 & 1 & 0\end{bmatrix}$$

And for a_3 and d_3:

$$a_3 = (\downarrow 2)[\frac{1}{\sqrt{4}}\begin{bmatrix}10 & 11 & 33 & 32\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}]$$

$$a_3 = \frac{1}{\sqrt{8}}\begin{bmatrix}21 & 65\end{bmatrix}$$

$$d_3 = (\downarrow 2)[\frac{1}{\sqrt{4}}\begin{bmatrix}10 & 11 & 33 & 32\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\end{bmatrix}]$$

$$d_3 = \frac{1}{\sqrt{8}}\begin{bmatrix}-1 & 1\end{bmatrix}$$

Now we have decomposed signal a_0 into three consecutive levels.

Continuation: Example of a Perfect Reconstruction Haar Wavelet Filter bank

If we then want to reconstruct the signal, we start at the highest decomposition level and work our way back. First, we upscale a_3 and d_3 by inserting zeros and then, we convolute with the reconstruction filters h and g respectively. We obtain a_2 by summing the upscaled and convoluted a_3 and d_3 coefficients. Recall that we are convoluting with the inverse of the filters, hence:

$$h_m^{inv} = h_{-m} = \frac{1}{\sqrt{2}}[1 \qquad 1]; \tag{51}$$

$$g_m^{inv} = g_{-m} = \frac{1}{\sqrt{2}}[-1 \qquad 1]; \tag{52}$$

As such a_2 can be calculated:

$$\tilde{a}_3 = \frac{1}{\sqrt{8}}\begin{bmatrix}0 & 21 & 0 & 65\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}$$

$$\tilde{a}_3 = \frac{1}{\sqrt{16}}\begin{bmatrix}0 & 21 & 21 & 65 & 65\end{bmatrix}$$

$$\tilde{d}_3 = \frac{1}{\sqrt{8}}\begin{bmatrix}0 & -1 & 0 & 1\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1\end{bmatrix}$$

$$\tilde{d}_3 = \frac{1}{\sqrt{16}}\begin{bmatrix}0 & \text{-1} & 1 & 1 & \text{-1}\end{bmatrix}$$

$$\gamma_2 = \tilde{a}_3 + \tilde{d}_3 = \frac{1}{\sqrt{16}}\begin{bmatrix}0 & 20 & 22 & 66 & 64\end{bmatrix}$$

$$\gamma_2 = \tilde{a}_3 + \tilde{d}_3 = \frac{1}{\sqrt{4}}\begin{bmatrix}0 & 10 & 11 & 33 & 32\end{bmatrix}$$

Note that the reconstructed signal is shifted, therefore, in order to get perfect reconstruction γ_2 has to be shifted:[51]

$$a_2[m] = \gamma_2[m-1]$$

$$a_2 = \frac{1}{\sqrt{4}}\begin{bmatrix}10 & 11 & 33 & 32\end{bmatrix}$$

Continuation: Example of a Perfect Reconstruction Haar Wavelet Filter bank

Similarly we obtain a_1 from summing upscaled and convoluted a_2 and d_2:

$$\tilde{a}_2 = \frac{1}{\sqrt{4}}\begin{bmatrix}0 & 10 & 0 & 11 & 0 & 33 & 0 & 32\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}$$

$$\tilde{a}_2 = \frac{1}{\sqrt{8}}\begin{bmatrix}0 & 10 & 10 & 11 & 11 & 33 & 33 & 32 & 32\end{bmatrix}$$

$$\tilde{d}_2 = \frac{1}{\sqrt{4}}\begin{bmatrix}0 & 2 & 0 & 1 & 0 & 1 & 0 & 0\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1\end{bmatrix}$$

$$\tilde{d}_2 = \frac{1}{\sqrt{8}}\begin{bmatrix}0 & 2 & -2 & 1 & -1 & 1 & -1 & 0 & 0\end{bmatrix}$$

$$a_1[m] = \gamma_1[m-1] = \tilde{a}_2 + \tilde{d}_2$$

$$a_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}6 & 4 & 6 & 5 & 17 & 16 & 16 & 16\end{bmatrix}$$

Finally, the - perfectly reconstructed - signal can be obtained:

$$\tilde{a}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 6 & 0 & 4 & 0 & 6 & 0 & 5 & 0 & 17 & 0 & 16 & 0 & 16 & 0 & 16\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}$$

$$\tilde{a}_1 = \frac{1}{\sqrt{4}}\begin{bmatrix}0 & 6 & 6 & 4 & 4 & 6 & 6 & 5 & 5 & 17 & 17 & 16 & 16 & 16 & 16 & 16 & 16\end{bmatrix}$$

$$\tilde{d}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 2 & 0 & 2 & 0 & 2 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 2 & 0 & 0\end{bmatrix} * \frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1\end{bmatrix}$$

$$\tilde{d}_1 = \frac{1}{\sqrt{4}}\begin{bmatrix}0 & 2 & -2 & 2 & -2 & 2 & -2 & 1 & -1 & 1 & -1 & 2 & -2 & 2 & -2 & 0 & 0\end{bmatrix}$$

$$a_0[m] = \gamma_0[m-1] = \tilde{a}_1 + \tilde{d}_1$$

$$a_0 = \frac{1}{\sqrt{4}}\begin{bmatrix}8 & 4 & 6 & 2 & 8 & 4 & 6 & 4 & 18 & 16 & 18 & 14 & 18 & 14 & 16 & 16\end{bmatrix}$$

$$a_0 = \begin{bmatrix}4 & 2 & 3 & 1 & 4 & 2 & 3 & 2 & 9 & 8 & 9 & 7 & 9 & 7 & 8 & 8\end{bmatrix}$$

3.2.2 Thresholding

When a signal has been decomposed in approximation and detail coefficients, the noise can be removed by thresholding certain detail coefficients. As mentioned, the detail coefficients represent the deviation of the signal from the approximation coefficients. Large detail coefficients may indicate a sudden intensity change in the actual signal, while small detail coefficients are more likely to be the result of small noise variation. Based on this wavelet coefficient property, a thresholding method was proposed by Donoho and Johnstone,[42] who opted to threshold a signal in its wavelet-decomposed state. The threshold can, furthermore, be determined from the data itself, which makes wavelet denoising well-suited for MA-XRF data were often no *a priori* information is known.

There are various methods to determine the threshold value for the wavelet coefficients, such as universal thresholding,[42] SURE (Stein's Unbiased Risk Estimate) thresholding,[52] Minimax thresholding,[53] and Bayesian thresholding.[54] These different thresholding methods are each optimal for specific cases.

The universal threshold is taken as an example, being the most widely used,[55] to explain the principle of thresholding. The formula for the universal threshold is as follows:

$$\lambda = \sigma\sqrt{2ln(N)}. \tag{53}$$

Here, σ is the average variance of the noise and N is the signal length. The value of the average variance σ is determined by the following equation:

$$\sigma = \frac{\text{Median}(|W_j|)}{0.6745}, \tag{54}$$

where W_j represents the detail coefficients at level j.

Once the thresholding value λ has been established, there are two often-used procedures to threshold the detail coefficients, called hard thresholding and soft thresholding. In hard thresholding, the detail coefficients whose absolute values are below the value of the threshold λ are set to zero, while the detail coefficients whose absolute value is above the threshold λ remain unchanged as follows:

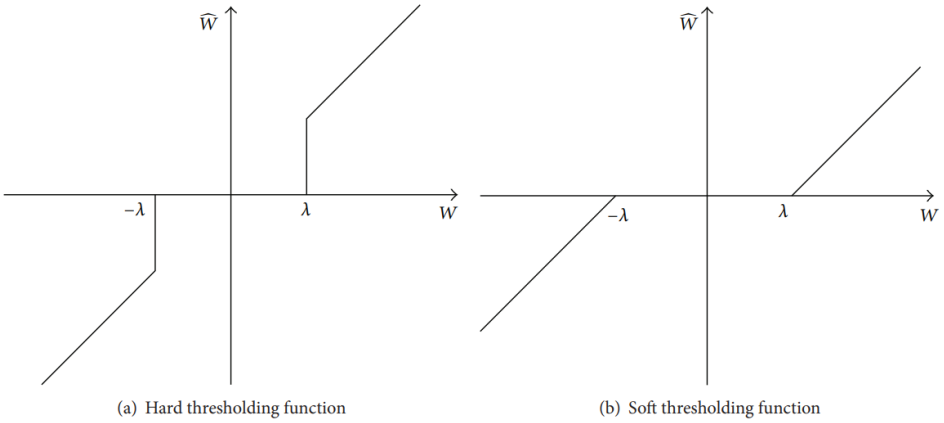
$$\widehat{W}_{j,k} = \begin{cases} W_{j,k} & \text{if} \quad |W_{j,k}| \geq \lambda \\ 0 & \text{if} \quad |W_{j,k}| < \lambda \end{cases}. \tag{55}$$

Here, the subscript k indicates each individual detail coefficient. For soft thresholding, the values of the detail coefficients for which their absolute value is lower than the threshold are similarly set to zero. The values of the detail coefficients, for which their absolute value is higher than the threshold value, are shrunk towards

zero by the same amount as the threshold value. Therefore, the soft thresholding function can be denoted as follows:

$$\widehat{W}_{j,k} = \begin{cases} sgn(W_{j,k})(|W_{j,k}| - \lambda) & \text{if} \quad |W_{j,k}| \geq \lambda \\ 0 & \text{if} \quad |W_{j,k}| < \lambda \end{cases} \tag{56}$$

Equation 55 and Equation 56 are visualised in Figure 13.



(a) Hard thresholding function (b) Soft thresholding function

Figure 13: (a) Hard thresholding and (b) soft thresholding.[55]

3.2.3 2D-Signal Denoising

Until now, the basics of wavelet filter bank denoising have been explained for a one-dimensional signal. However, for the purpose of spatially denoising MA-XRF data, the method of wavelet filter bank denoising will be applied to a two-dimensional signal. Therefore, in this section wavelet filter bank denoising for a two-dimensional signal is explained.

For two-dimensional signals, such as images, the tree-like structure of wavelet filter banks as shown in Figure 8 is expanded. Instead of just one set of detail coefficients, there are three sets of detail coefficients per decomposition level, called horizontal, vertical, and diagonal detail coefficients, see Figure 14a. As their name implies, the horizontal detail coefficients capture horizontal details, the vertical detail coefficients capture vertical details, and the diagonal detail coefficients capture diagonal details.

To obtain all these coefficients, the signal is filtered two consecutive times instead of one time per level. First, the rows are filtered, in which each row can be seen as a one-dimensional signal. After filtering, the rows are each decimated by

a factor of two, hence, the number of columns becomes twice as small. Hereafter, the columns are filtered and consequently decimated, reducing the number of rows by a factor of two as well, hence, the ratio between rows and columns stays equal. The process of the two-dimensional signal decomposition can be seen in Figure 14.

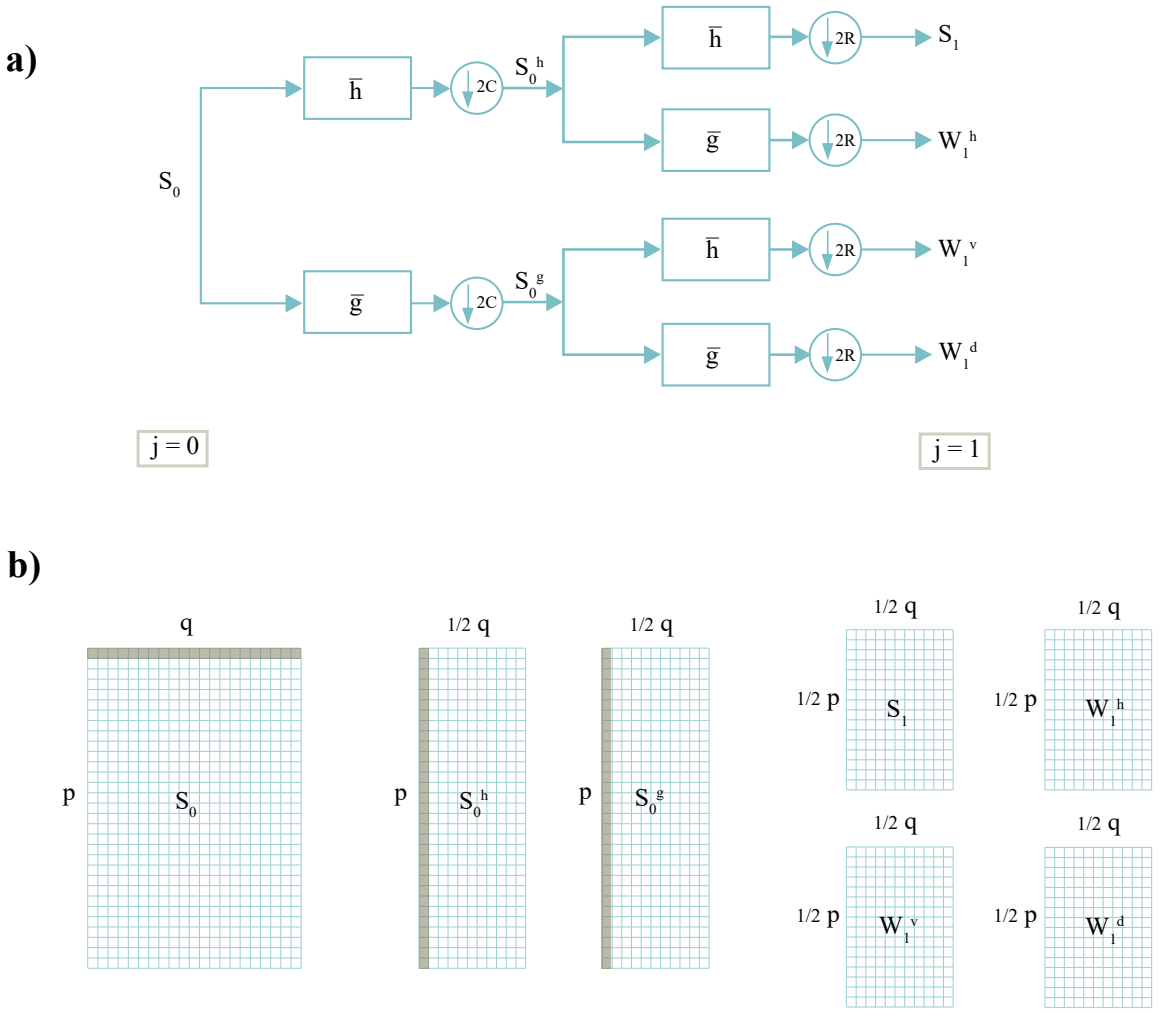


Figure 14: a) 2D analysis filter bank of 1 decomposition level. b) Visualisation of the data geometry during 2D signal decomposition. S_0 is the two-dimensional input signal and the decomposition level is denoted by j. $\bar{h}$ and $\bar{g}$ are the low-pass and high-pass decomposition filters, respectively. $2C$ implies the downsizing of the columns by a factor of two, similarly, $2R$ implies the downsizing of the rows by a factor of two. S_0^h and S_0^g represent the row-filtered signal of the low-pass filter and the high-pass filter, respectively. S_1 represent the approximation coefficients, W_1^h represents the horizontal detail coefficients, W_1^v represents the vertical detail coefficients, and W_1^d represents the diagonal detail coefficients. Furthermore, p denotes the number of rows and q the number of columns of the input signal.

For deriving the formulas of the approximation coefficients and the three sets of detail coefficients, we recall Equation 44 and Equation 43. The first filter bank step results in S_0^h and S_0^g:

$$S_{0,p}^h = \sum_{m\in\mathbb{Z}} \overline{h_{m-2k}} S_{0,m,p} \tag{57}$$

$$S_{0,p}^g = \sum_{m\in\mathbb{Z}} \overline{g_{m-2k}} S_{0,m,p} \tag{58}$$

Here, p is the row number. The approximation coefficients are obtained by filtering the two-dimensional signal both horizontally and vertically with low-pass filter h:

$$S_{1,q} = \sum_{m\in\mathbb{Z}} \overline{h_{m-2k}} S_{0,m,q}^h \tag{59}$$

Here, q is the column number. The horizontal detail coefficients are obtained by filtering the 2D-signal horizontally with a low-pass h filter and vertically with a high-pass filter g:

$$W_{1,q}^h = \sum_{m\in\mathbb{Z}} \overline{g_{m-2k}} S_{0,m,q}^h \tag{60}$$

The vertical detail coefficients are obtained by filtering the signal horizontally with a high-pass g filter and vertically with a low-pass filter h:

$$W_{1,q}^v = \sum_{m\in\mathbb{Z}} \overline{h_{m-2k}} S_{0,m,q}^g \tag{61}$$

The diagonal detail coefficients are obtained by filtering the signal horizontally and vertically with a high pass filter g:

$$W_{1,q}^d = \sum_{m\in\mathbb{Z}} \overline{g_{m-2k}} S_{0,m,q}^g \tag{62}$$

After the 2D-signal is decomposed, the detail coefficients are thresholded as explained in section 3.2.2 and subsequently the signal can be reconstructed.

3.3 Wavelet Filter bank Denoising: a Procedure for Spatially Denoising MA-XRF Data

In the previous sections, the reason for choosing wavelet denoising for MA-XRF data denoising was justified and the working principle of wavelet filter banks was explained. In this section, that knowledge is used for design considerations in order to create a wavelet denoising procedure that is suitable for MA-XRF data denoising.

3.3.1 Choosing an Appropriate Wavelet

Not all wavelets can be used for perfect reconstruction filter bank denoising. It is necessary for the wavelets to be able to construct an orthonormal basis and for their transform to have compact support. The latter is achieved when the transform has zero values outside a bounded interval in the frequency domain.[46] This leaves us the Haar wavelet, the Daubechies wavelet family, Symlets, and Coiflets.

Still, within these wavelet families, there are large differences between wavelets as well; for example, their number of vanishing moments. A wavelet ψ has p vanishing moments if:

$$\int_{-\infty}^{\infty} \psi(x) x^k dx = 0 \qquad \text{for} \quad 0 \leq k < p. \tag{63}$$

Polynomials having a degree from 0 until $p-1$ can be accurately represented by the scaling functions. The higher the number of vanishing moments, the more complex signals can be represented, or, the more small detail coefficients are obtained, making a sparser representation of the signal possible.

The factors that influence the shape of ψ are, besides its number of vanishing moments, the support size and regularity.[46] Where regularity is a measure of the smoothness of the wavelet function. The more vanishing moments, the higher the regularity. Which wavelet ψ is best suited for denoising, depends on the signal. For example, if the signal is mainly regular (smooth) and has but a few isolated singularities, a wavelet with many vanishing points is desired since it will generate a lot of small detail coefficients. If singularities are close to one another, a wavelet with a small support is optimal. However, it is not possible to have both a small support and many vanishing moments for there is a trade-off between the two.

The regularity of ψ does not influence the amplitude of the detail coefficients in the way the support size and amount of vanishing moments do, the regularity is mainly an issue for the visual appeal of the denoised signal. Namely, when a signal is reconstructed after thresholding, the error which is added to a detail coefficient will result in a smooth error if the wavelet function is smooth, and in an irregular error if the wavelet function is irregular. The reconstructed signal will consequently appear smooth if the added error is smooth and vice versa, even though the errors are of similar magnitude.[46] Therefore, images are often denoised using wavelets with high regularity.

Since the objective of denoising the MA-XRF data is more focused on improving the process of evaluating spectral data than obtaining aesthetically pleasing peak intensity maps, regularity is not a factor of great importance in our case. What is important is the ability to denoise while refraining from smoothing out important features within the data. Since the spatial MA-XRF signal is characterised by quite some intensity fluctuations, a relatively low number of vanishing moments

is required. Still, considering there is no exact method for pre-determining the optimal number of vanishing moments, a subset of the wavelet families stated above is going to be tested for their performance in the denoising of MA-XRF data. The wavelets chosen for testing the perfect reconstruction filter bank are the Haar, Db2-Db6, Sym2-Sym6, and Coif1-Coif2.

3.3.2 Choosing Thresholding Method

Since there are various wavelets that each have their own advantages for denoising different signal types, there are also various thresholding selection procedures, as has been discussed in section 3.2.2.

Choosing the right threshold selection procedure is critical for correctly denoising MA-XRF data. Signals can be distorted when a threshold value is too high. However, data can remain polluted with noise when a threshold value is too low.

In the case of MA-XRF, the threshold selection procedure is, furthermore, bound to some constraints, as was stated in section 3.1.1. In the first place, it is essential that no prior information is needed for determining the threshold and, next to that, it is important that the absolute value of the signal intensity is not substantially compromised. Finally, the goal is to eliminate shot noise, while most methods for determining the threshold value assume Gaussian noise.[56]

This set of constraints limits the amount of possible threshold selection procedures significantly. For example, thresholding methods that do account for shot noise are often not purely data-driven or need processing in the image domain, which nullifies the wavelet property of maintaining the absolute values of the data intact.

Luisier *et al.*[57] described a data-driven threshold selection procedure for shot noise-polluted data, called PURE-LET. However, in order to calculate the Mean-Squared-Error (MSE), used to optimize the threshold, still a statistical estimate has to be constructed beforehand. This is fine when dealing with medical images, having large amounts of similar samples, but in the case of MA-XRF data, this is not ideal.

Being the most widely used thresholding function and being both data-driven and able to maintain the absolute value of the data, universal thresholding has been used as our base procedure for determining the threshold values. Unlike the common use of the universal threshold, the threshold is determined for each level instead of taking one value for every level. Furthermore, in order to maintain the absolute value of the data, hard thresholding is applied.

Nevertheless, this procedure is insufficient for our purposes, and it fails to account for shot noise and has no awareness of MA-XRF's spectral domain. In our work, we prepare a solution for this problem by exploiting the 3-dimensional

nature of MA-XRF datasets. To accomplish this, Equation 54 is expanded by calculating the variance σ not using the detail coefficients of one channel, but using its neighbouring channels as well, as can be seen in Figure 15. Recall Equation 54:

$$\sigma = \frac{Median(|W_j|)}{0.6745}$$

Then, the $\sigma_{j,i}$ for the detail coefficients of channel i is determined for the averaged sum Ω of the detail coefficients at channel i and its neighbouring channels as follows:

$$\Omega_{j,i} = \frac{\sum_{i-s}^{i+s} W_{j,i}}{2s+1} \tag{64}$$

$$\sigma_{j,i} = \frac{Median(|\Omega_{j,i}|)}{0.6745} \tag{65}$$

where s stands for the number of neighbouring channels on the left and on the right of channel i. The advantage of using the average detail coefficients of a channel and its neighbouring channels to base the standard deviation on is the fact that the effect of shot noise is reduced. By averaging the detail coefficients, large intensity fluctuations due to shot noise are largely minimized. This results in the standard deviation being lower, hence, the threshold being lower.

When one bases the standard deviation on one channel alone, the threshold would be too high and too many detail coefficients would be set to zero. Therefore, it is necessary to lower the standard deviation. Moreover, by lowering the standard deviation on a per-channel basis, we automatically set higher thresholds at higher-intensity peaks and lower thresholds at lower-intensity peaks. This ensures the removal of noise at high-intensity peaks while retaining the signal of low-intensity peaks, even though the noise of the high-intensity peak might be of a higher magnitude than the signal of the low-intensity peak.

The optimal number s for MA-XRF wavelet denoising is unknown, but can only be found through testing. An initial estimation, however, allows the quicker evaluation of other filter design factors. Then, after a selection of well-performing filters has been made, the optimal value of s can be determined.

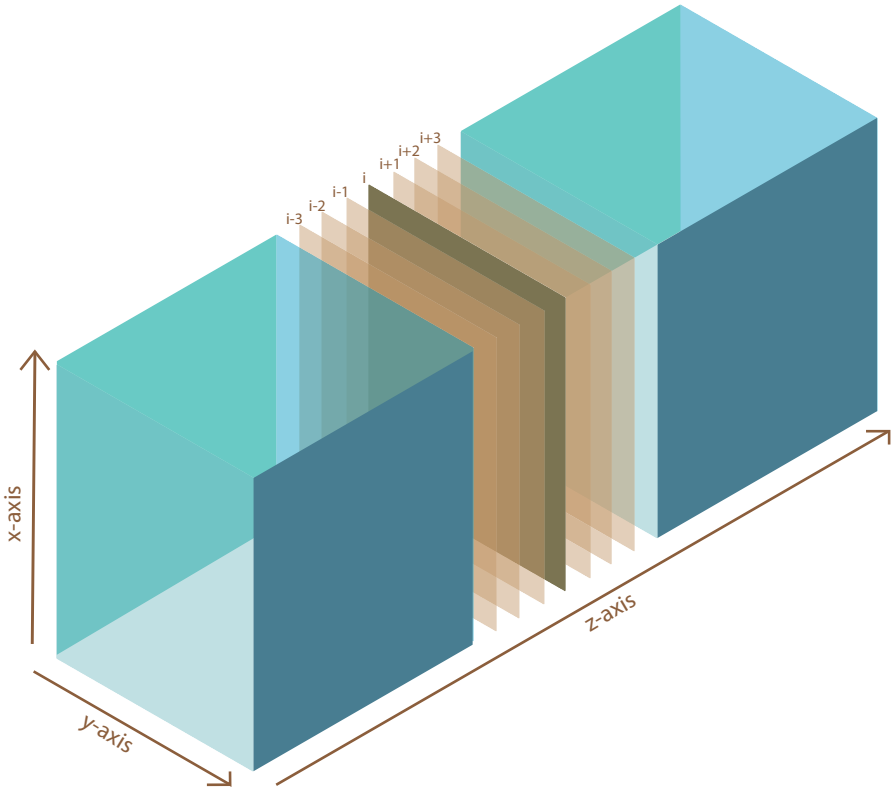


Figure 15: Schematic overview of neighbouring channels of channel i within a hyperspectral datacube. The x and y axes represent the spatial pixel position, while the z axis represents the channels. Here the case of $s = 3$, i.e. 3 neighbouring channels on each side, is depicted.

3.3.3 Choosing the Number of Decomposition Levels

The amount of denoising of the dataset is not only dependent on the threshold, but also on the resolution in which the data is thresholded. Every increasing level j, the resolution becomes coarser. Hence, at lower j, fluctuations of the detail coefficients are more likely noise inflicted, while at higher j fluctuations of the detail coefficients are more likely to include signal features as well. It is important to choose the right amount of decomposition levels where the noise is thresholded and the important features are not. There is, however, no clear procedure to predetermine the optimal amount of decomposition levels. Therefore, just like with the different mother wavelets, the optimal amount of decomposition levels for denoising the MA-XRF data is going to be tested. The amount of decomposition levels that is chosen to be tested is 6, scaling the mother wavelets until a factor $2^6 = 64$.

4 Method and Case study

In this chapter, the method for testing the wavelet filter described in section 3.3 is outlined. The different filter design combinations are explained and the quality measures that indicate the performance of respective filters are introduced. Thereafter, a case study of a daguerreotype, on which several filters will be tested, is specified.

4.1 Method

Wavelet filter bank denoising is proposed as a method suitable to denoise MA-XRF data. Spatially denoising MA-XRF data has not yet become standard practice, hence, there are few methods to compare this new method with. Yet, as mentioned in section 2.1.3, Kogou *et al.* used a 3 by 3 kernel for spatial median filtering of the data. Therefore, the proposed wavelet filter bank denoising method will be compared to a median filtering algorithm. Furthermore, the filter bank design described in section 3.3 left some options open with respect to what mother wavelet suits best, what level of decomposition should be applied, and how many neighbouring channels should be used to determine σ. Accordingly, several different combinations for wavelet filter bank design options will be tested. As was mentioned in section 3.3.2, an initial estimation of s allows us to first optimize for the mother wavelet and the number of levels. Then, with a specific selection of filters, we optimize for s.

The evaluation of the different filters is done by means of different quality measures: Signal-to-Noise Ratio (SNR), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) of the filtered noisy data with respect to ground truth data. Both quality measures of individual spectra and more broad quality measures of summed spectra are determined. Moreover, a visual inspection of the error distribution images, pixel spectra, sum spectra, and peak intensity maps will be done for the various filters. Finally, every filter will be evaluated for various simulated dwell times, which correspond to various noise levels.

4.1.1 Filter Designs

A wavelet filter bank consists of a lot of different design parameters. These parameters were described in section 3 and argumentation for the choice of certain parameters was given in section 3.3. The choice of the design parameters was narrowed down based on literature indicating that certain properties would be suitable for MA-XRF data denoising. The final choice for the parameters was not as self-evident and, therefore, these parameter choices are instead determined based on quality indicators, namely SNR, MAE, and RMSE. However, not all optimizable

parameters are tested, so has the threshold method been chosen. The optimizable parameters used for the wavelet filter bank and whether they are optimized or adopted are indicated in Table 1.

Here it can be seen that the preliminary estimation for the number of neighbouring channels is 5 channels on each side. To obtain this initial estimation for s, testing was done with a 3-level db3 decomposition. The different performance of the filter using $s = 0$. $s = 5$, and $s = 10$ neighbouring channels were analysed, and the best performance was achieved with $s = 5$.

Table 1: Optimizable Design Parameters

	Parameter	Specification
Optimized	Mother Wavelet	Haar, db(2-6), sym(2-6), coif(1,2)
Optimized	Decomposition levels	level 1 until 6
Estimated/ Optimized	Number of Neighbouring Channels s	Preliminary Assumption: $s = 5$
Adopted	Threshold	Universal Threshold

4.1.2 Quality Measures

For assessing the quality of the filters, the SNR, RMSE, and MAE of the filtered signal will be determined using a synthetic dataset, where the ground truth signal is available.

Signal-to-Noise Ratio (SNR)
The signal-to-noise ratio (SNR) is an indicator that characterizes the quality of noise suppression during filtering.[58] At the beginning of this research, we stated that the signal-to-noise ratio should become optimized for lower dwell times. Therefore, to determine which method for spatially denoising MA-XRF data is best suited towards this objective, the SNR of the filtered data is calculated.

$$SNR = \frac{\sum_{i=1}^{N} s_i}{\sum_{i=1}^{N} |x_i - s_i|} \tag{66}$$

Here, x_i is the filtered data consisting of noise and signal, and in the filtered case hopefully less noise than originally. s_i is the ground truth signal. s and x will be vectors when calculating the spectral SNR and matrices when calculating the spatial SNR, i.e., s_i is the ground truth value of channel i when calculating the spectral SNR and s_i is the ground truth value of pixel i when calculating the spatial SNR.

Root Mean Square Error (RMSE)
The RMSE is calculated to determine how large the error is between the filtered data and the ground truth. The larger the deviation from the ground truth, the larger the error is penalized (quadratically).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} ||x_i - s_i||^2}{N}} \tag{67}$$

Here, x_i is the filtered signal, s_i is the ground truth, and N is the amount of values in the signal.

Mean Absolute Error (MAE)
Through the proportionally larger weighting of large deviations from the ground truth, the RMSE is biased towards elements with high photon counts. The MAE uses a weighting that is linearly proportional to the error rather than quadratically and, therefore, does not have that problem.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |x_i - s_i| \tag{68}$$

Here, x_i is the filtered signal, s_i is the ground truth, and N is the amount of values in the signal.

Together, for our synthetic dataset, these three quality measures give an accurate indication of the suitability of a certain filter for denoising MA-XRF data. Ideally, both RMSE, and MAE are low, since this would indicate a higher resemblance between the ground truth and the filtered data. Meanwhile, it is important that the SNR is high, for that indicates how resistant the signal is to noise.

4.1.3 Synthetic Datasets

To properly assess filters with the quality measures described in section 4.1.2, a ground truth signal is necessary. To this end, we created a synthetic dataset. The synthetic dataset was created by assigning real-world MA-XRF pixel spectra to the spatial pattern covering a 780×780 binary image retrieved from a 1951 USAF resolution test chart of 96 dpi, as shown in Figure 16. These spectra were obtained from an actual MA-XRF scan of a daguerreotype, which will be elaborated upon in section 4.2.1. This scan was made with a dwell time of 100 milliseconds and a step size of 50 micrometres. In doing so, a scan of 376 by 260 pixels and 4096 channels was obtained. For now, it is important to clarify that the brighter areas in the daguerreotype are rich in mercury and the darker areas are poor in mercury. Hence, the mercury-rich pixel spectrum was created by taking the sum average of a

mercury-rich area of the daguerreotype, indicated by the blue squares in Figure 17. Similarly, the mercury-poor pixel spectrum was created by taking the sum average of a mercury-poor area, indicated by the red square in Figure 17. The mercury-rich and mercury-poor pixel spectra are plotted in Figure 18 and the difference between the spectra is plotted in Figure 19. The green bar represents the area of the Hg L3 peak. In the synthetic dataset, a mercury-poor pixel spectrum was assigned to the darker areas, while a mercury-rich pixel spectrum was assigned to the brigher areas of the 1951 USAF Resolution Test chart. A peak intensity map of the Hg L3 peak of the synthetic dataset is shown in Figure 20.

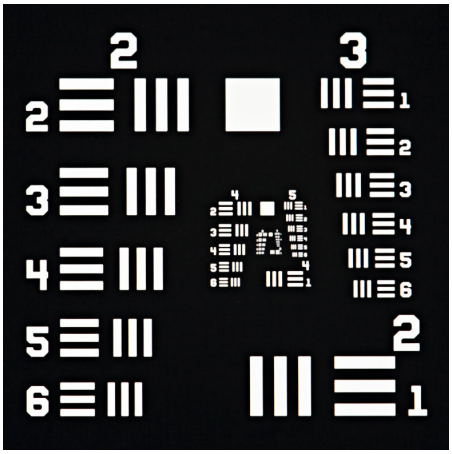


Figure 16: 1951 USAF Resolution Test Chart used for creating the spatial distribution of mercury for the synthetic dataset by assigning MA-XRF spectra to it

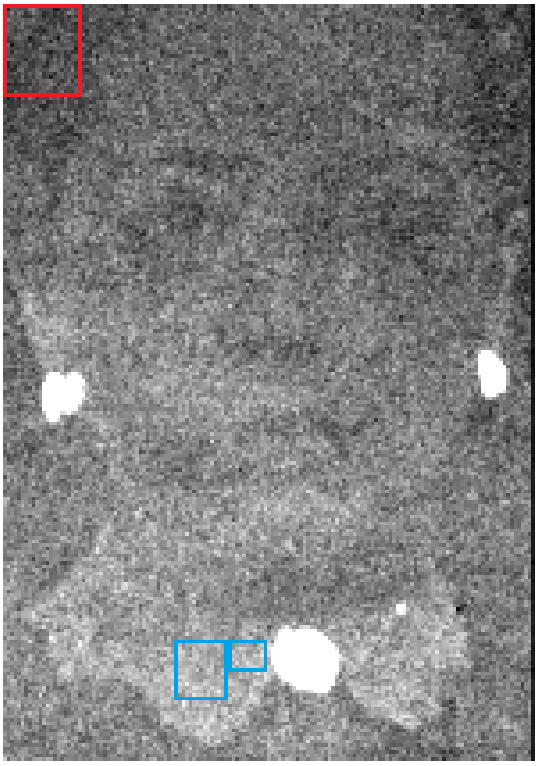


Figure 17: Peak intensity map of the Hg L3 peak of a daguerreotype. In a peak intensity map, the relative intensity of photon counts per fluorescence line per pixel is visualised. This is done by summing several channels covering the peak created by the fluorescence line of an element. The brighter areas represent pixels with more Hg L3 photon counts, while the darker areas represent pixels with less Hg L3 photon counts. Hence, the blue areas are rich in mercury and the red area is poor in mercury

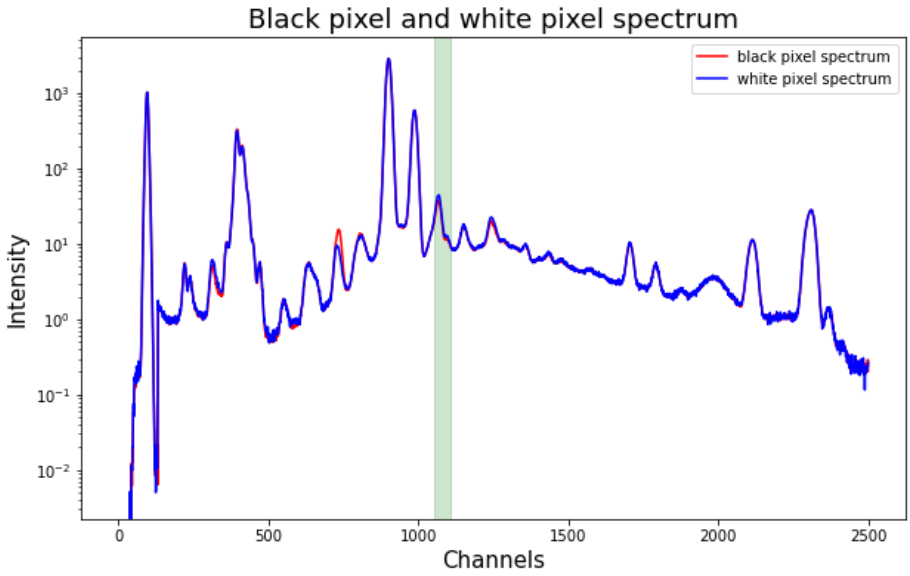


Figure 18: Average pixel spectra of the black area and white area pointed out by the red and blue boxes respectively in Figure 17

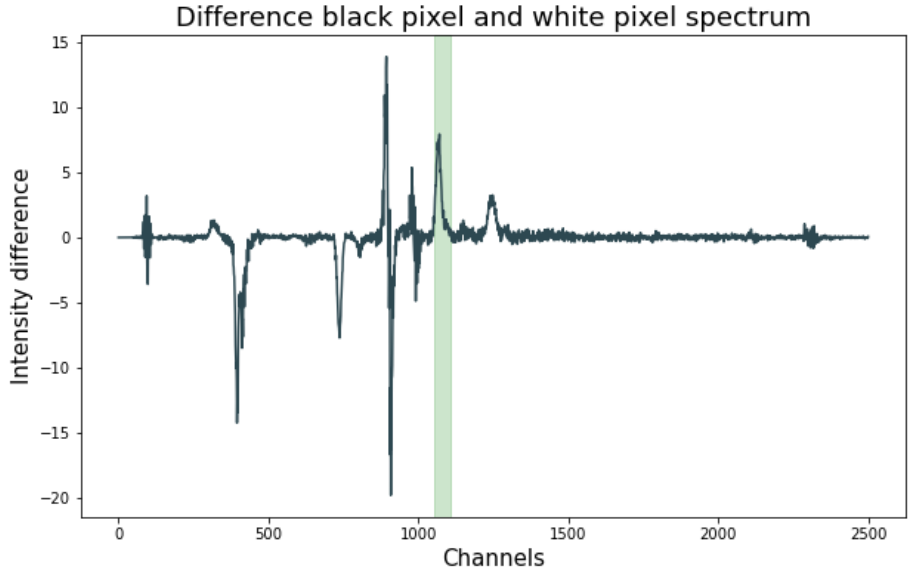


Figure 19: Difference between the average spectra of the black and white areas

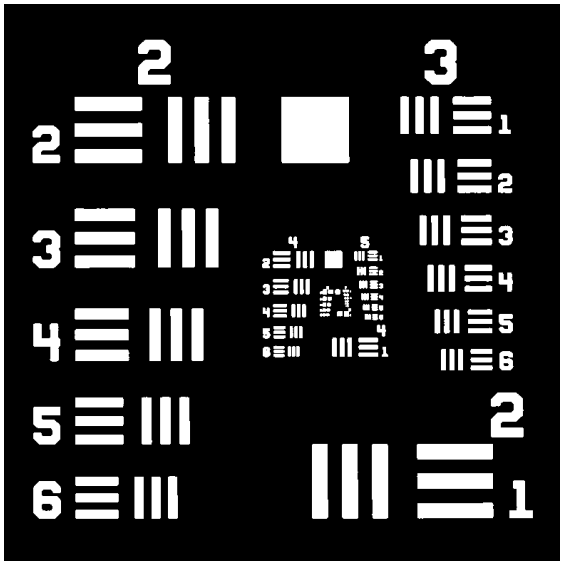


Figure 20: Hg L3 peak intensity map of the synthetic dataset. Created by summing the channels covering the Hg L3 peak.

From the original clean synthetic dataset, other datasets have been derived each with increasingly more noise. This has been done by artificially reducing the dwell time per pixel, which comes down to reducing the photon counts within the pixel spectra simply by multiplying it by 0.75 and 0.5 to obtain a dwell time of 0.75 s and 0.5 s, respectively. After the reduction of the dwell time, shot noise is applied, using the *numpy.random.poisson* command in Python. Due to the lesser amount of photons, shot noise becomes more prominent at low dwell times. In this manner, three datasets have been obtained having varying noise levels. These three datasets, shown in Table 2, are used to test how well the filters described in section 4.1.1 perform for different degrees of noise. The noisy Hg L3 peak intensity maps of the three noise-perturbed datasets are shown in Figure 21.

Table 2: Testing Data

Test Nr.	Dwell times
T. 1	1.0 s
T. 2	0.75 s
T. 3	0.5 s

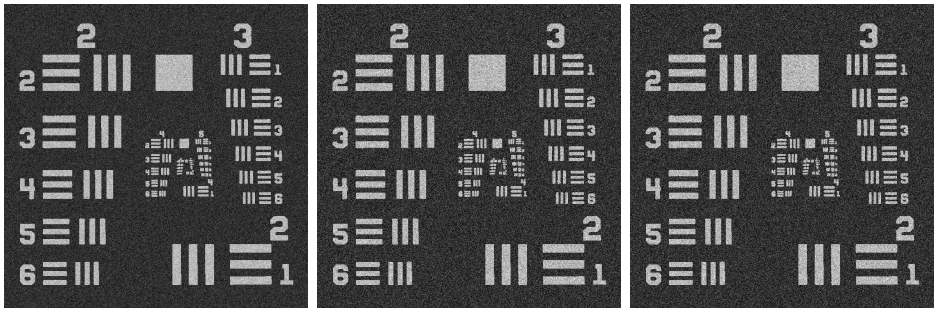


Figure 21: Hg L3 peak intensity maps of a synthetic MA-XRF dataset perturbed with different levels of noise, Left: Poisson noise at a dwell time of 1 sec. Middle: Poisson noise at a dwell time of 0.75 sec. Right: Poisson noise at a dwell time of 0.5 sec.

4.1.4 Tests

The synthetic dataset is used to determine the performance quality of wavelet filters for different parameters. In order to prevent large homogeneous areas from influencing the outcome of the test, the performance quality is not determined for the entire dataset, but on a subset of the data which is indicated by the blue square in Figure 22. By doing so, rewarding performance on small detailed variation, while penalizing over-smoothing.

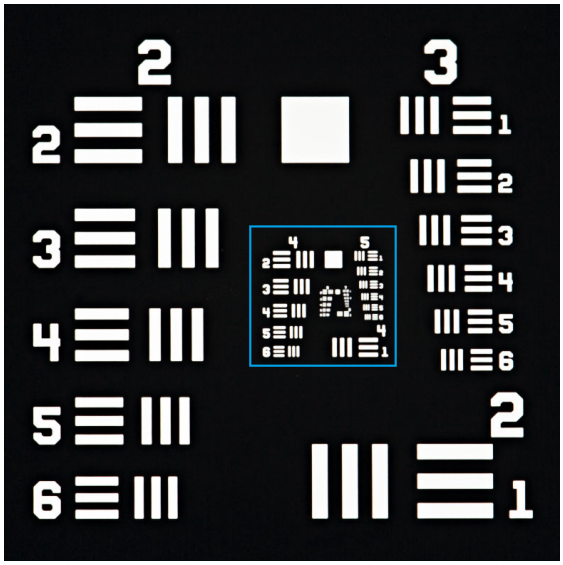


Figure 22: Subset of the Hg L3 peak intensity map of the synthetic dataset indicated by the blue square

First Testing Round

As mentioned, the amount of neighbouring channels for determining σ is initially set at $s = 5$. Hence, in the first testround, only the effect of the mother wavelet and the number of decomposition levels will be evaluated, using SNR, MAE, and RMSE as performance quality indicators.

The various options for the parameters *mother wavelet* and *decomposition level* result in 78 possible filter combinations. These filters are shown in Table 3.

Mother Wavelets \ Levels	1	2	3	4	5	6
Haar	F.1	F.2	F.3	F.4	F.5	F.6
DB2	F.7	F.8	F.9	F.10	F.11	F.12
DB3	F.13	F.14	F.15	F.16	F.17	F.18
DB4	F.19	F.20	F.21	F.22	F.23	F.24
DB5	F.25	F.26	F.27	F.28	F.29	F.30
DB6	F.31	F.32	F.33	F.34	F.35	F.36
SYM2	F.37	F.38	F.39	F.40	F.41	F.42
SYM3	F.43	F.44	F.45	F.46	F.47	F.48
SYM4	F.49	F.50	F.51	F.52	F.53	F.54
SYM5	F.55	F.56	F.57	F.58	F.59	F.60
SYM6	F.61	F.62	F.63	F.64	F.65	F.66
COIF1	F.67	F.68	F.69	F.70	F.71	F.72
COIF2	F.73	F.74	F.75	F.76	F.77	F.78

Table 3: Tested Filter Designs in First Round with $s = 5$

Second Testing Round

Using a selection of 9 filters obtained from the previous testing round, the effect of the number of neighbouring channels used for the universal threshold calculation, i.e. parameter s, will be studied. As mentioned, for obtaining the initial assumption of $s = 5$, a rough comparison between filters, having $s = 0$, $s = 5$, and $s = 10$, was made. Now, in this second testing round, for every value of s between 0 and 10, the filter quality is going to be assessed based on the same quality measures used in the first testing round, namely visual analysis, SNR, RMSE and MAE.

Third Testing Round

After having determined the optimal design parameters for the wavelet filter, and the optimal number of neighbouring channels, the filter will be compared to two median filters, one having a 3×3 kernel and the other having a 5×5 kernel. This evaluation will be done by applying the same quality measures as were used in the first and second testing round: SNR, RMSE, and MAE. Moreover, the performance quality of the filters will be compared for synthetic data subject to different degrees of noise as shown in Table 2.

4.2 Case study

In this section, a case study is described that validates the proposed wavelet filter for use on real-world (non-synthetic) MA-XRF data. The study considers a MA-XRF dataset, obtained from a daguerreotype, shown in Figure 23. First, the daguerreotype and its physical properties will be explained and, thereafter, the settings for obtaining the data and how the data is treated will be elaborated upon.

Figure 23: Left: Daguerreotype of a woman, from private collection. Right: MA-XRF scanned area.

4.2.1 The Daguerreotype

A daguerreotype is the result of a chemical photographing process invented by Louis Daguerre. Being introduced to the world in 1839, the daguerreotype became the first commercially available method of photography and was used till the early 1860s.[59]

A cross-section of a daguerreotype is shown in Figure 24. The base of a daguerreotype is formed by a copper (Cu) plate that is coated with a silver (Ag) layer and subsequently polished. The polished metal plate then becomes exposed to halogens (often iodide (I) is used), making it responsive to light exposure. At this point, it can be used to make a photograph. When the plate is exposed to light, salts are formed by the halogen reacting with the silver. The more light hits the plate, the more salt particles are formed, hence, the higher the particle density. To visualise the photograph, it has to be developed, which is done using mercury (Hg) fumes. During the developing process, silver particles amalgamate and silver-mercury nano-particles are formed. The photo then becomes visible through the property of mercury particles which, through their high density, reflect light diffusely. The places with high silver-mercury particle density appear as light, while those with low silver-mercury particle density appear as dark. As such, a grey-scale image is obtained.[60] In order to prevent the image from further developing due to extended light exposure, the remaining salts (silver-iodide) are then removed from the plate. Hereafter, the photo is often gilded with gold (Au).[61]

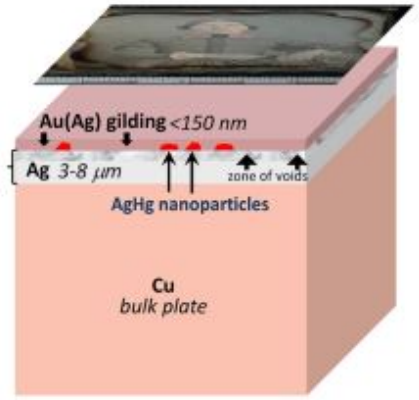


Figure 24: Anatomy of a daguerreotype.[61]

Several factors make the daguerreotype ideal as a case study to test the wavelet filters: in the first place, its simple and largely predictable material composition, and secondly, the homogeneous support material. Moreover, the mercury distribution being an exact match with the visual image allows for direct visual comparison between the elemental mercury/gold distribution map and the photograph. Testing the filter would come down to evaluating to what extent the visual image can be recovered from the MA-XRF data.

4.2.2 The MA-XRF Data of a Daguerreotype

The daguerreotype from Figure 23 has been scanned with a Bruker M6 Jetstream, using a dwell time of 30 milliseconds and a step size of 50 micrometres. A scan of 376 by 260 pixels and 4096 channels was obtained.

The first step of data treatment involves multiplying each spectrum by a factor $\frac{1000}{30}$ to simulate a dwelltime of 1 s. Also, the channels covering the zero peak are determined through visual inspection of the sum spectrum and the data is normalized as such that every zero peak has an intensity of 10,000. Herewith, intensity fluctuations due to dwell time variations are corrected, as described in section 2.1.3. Thereafter, the data is cropped from 4096 channels to 2500 channels in order to reduce the amount of data. Finally, the data is filtered with a selection of the tested wavelet filters and a median filter.

5 Results and Discussion

In this chapter, the results from evaluating the filters, as described in 4.1, are presented and discussed. We begin with addressing the results of the first testing round. After the first testing round, 9 wavelet filters are selected that are used in the second testing round. From the second testing round, the optimal amount of neighbouring channels to determine the standard deviation σ for the universal thresholding is obtained. Subsequently, a selection of optimal wavelet filter designs is made by combining the results from the first and second testing round. Hereafter, the performance quality of these designs is compared to that of two median filters for different amounts of noise. Then, the applicability of the wavelet filter is considered through filtering an actual MA-XRF dataset. The chapter ends with a general discussion on the applicability of the wavelet filter, taking the three testing rounds and the case study into account.

5.1 Testing Round 1

In the first testing round, a set of 78 different wavelet filters is evaluated. Here, the value of s is set as 5 and optimization is focussed on finding the optimal mother wavelet and decomposition level. In Figure 25, the performance quality (MAE, RMSE, and SNR) of these filters is visualised.

5.1.1 Observations Testing Round 1

Explaining and Observing the Performance Quality Graph. The performance quality plotted in Figure 25 is based on the entire dataset. The *spectral average (avg)* performance quality, refers to the performance quality of every individual spectrum. Here, the performance quality is first calculated per spectrum and then these values are averaged. This is similar to the *spatial average (avg)* performance quality, where the performance quality of every individual channel is calculated before averaging. With the *spectral sum* performance quality, first all the spectra are summed and, then, the performance quality is determined. Similarly, the *spatial sum* performance quality is determined.

The MAE, RMSE, and SNR are plotted against the number of levels for the 13 different mother wavelets: haar, db2, db3, db4, db5, db6, sym2, sym3, sym4, sym5, sym6, coif1, and coif2. The performance quality at $x = 0$, indicates the MAE, RMSE, and SNR of the unfiltered data.

Both spectral and spatial performance quality seem to increase with increasing decomposition level for every mother wavelet, However, most improvement is made in the first few levels, after which the graphs flatten. When looking at the spectral performance quality, particularly the haar, but also the sym2, db2, and

coif1 are performing very well. These 4 mother wavelets have the best spatial average performance quality as well. And although the haar wavelet has a high spatial sum performance quality, sym2, db2 and coif1 actually perform the worst.

Visualising Spatial Performance. The spatial performance quality is an indication of how well the filtered channels resemble the ground truth channels. Hence, the spatial similarity between the ground truth data and the filtered data can also be visualised by a peak intensity map. In Figure 26, the mercury L3 peak intensity distribution is mapped for the haar and the db2 filtered data. Despite their difference in spatial performance quality, here, not a clear distinction between the two can be seen. Remarkable, however, is the fact that for both wavelets, the peak intensity maps seem to be subject to artifacting at level 6.

Plotting the Average Spectral and Spatial Performance Quality of the Mercury L3 Peak. To verify how the spatial performance seemed to get worse after a certain level when regarding the peak intensity maps, while this did not appear to be the the case when regarding Figure 25, the average spectral and spatial performance quality of the mercury peak is investigated. So, instead of calculating the performance quality of the entire dataset, the performance quality of an area of 50 channels, which approximately covers the mercury peak, is determined for every pixel/channel and then averaged. In Figure 27 and Figure 28, it is clearly visible that for each wavelet the performance quality seems optimal at level 4.

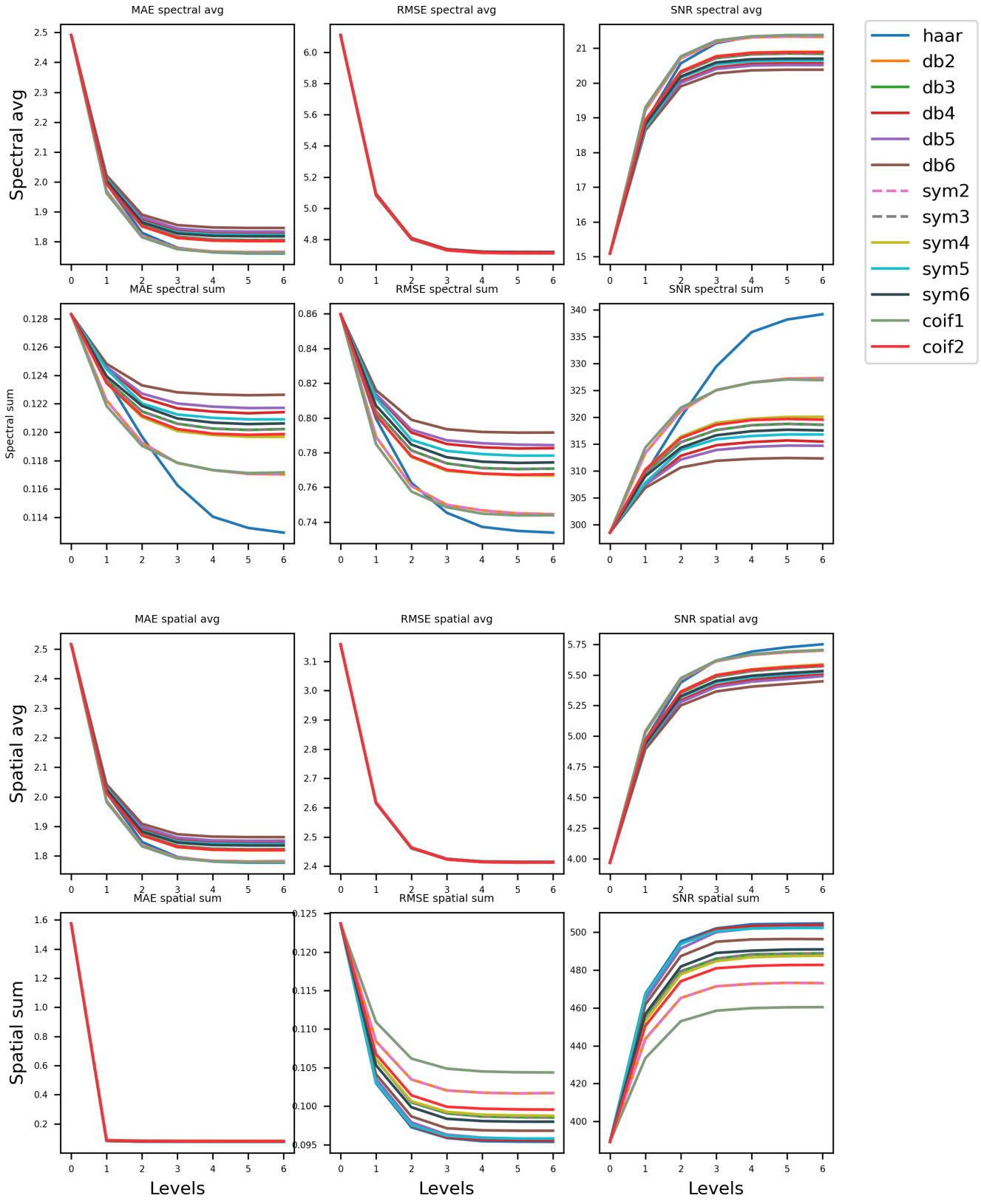


Figure 25: Filter performance, indicated by MAE, RMSE, and SNR for various mother wavelets, plotted against the number of decomposition levels

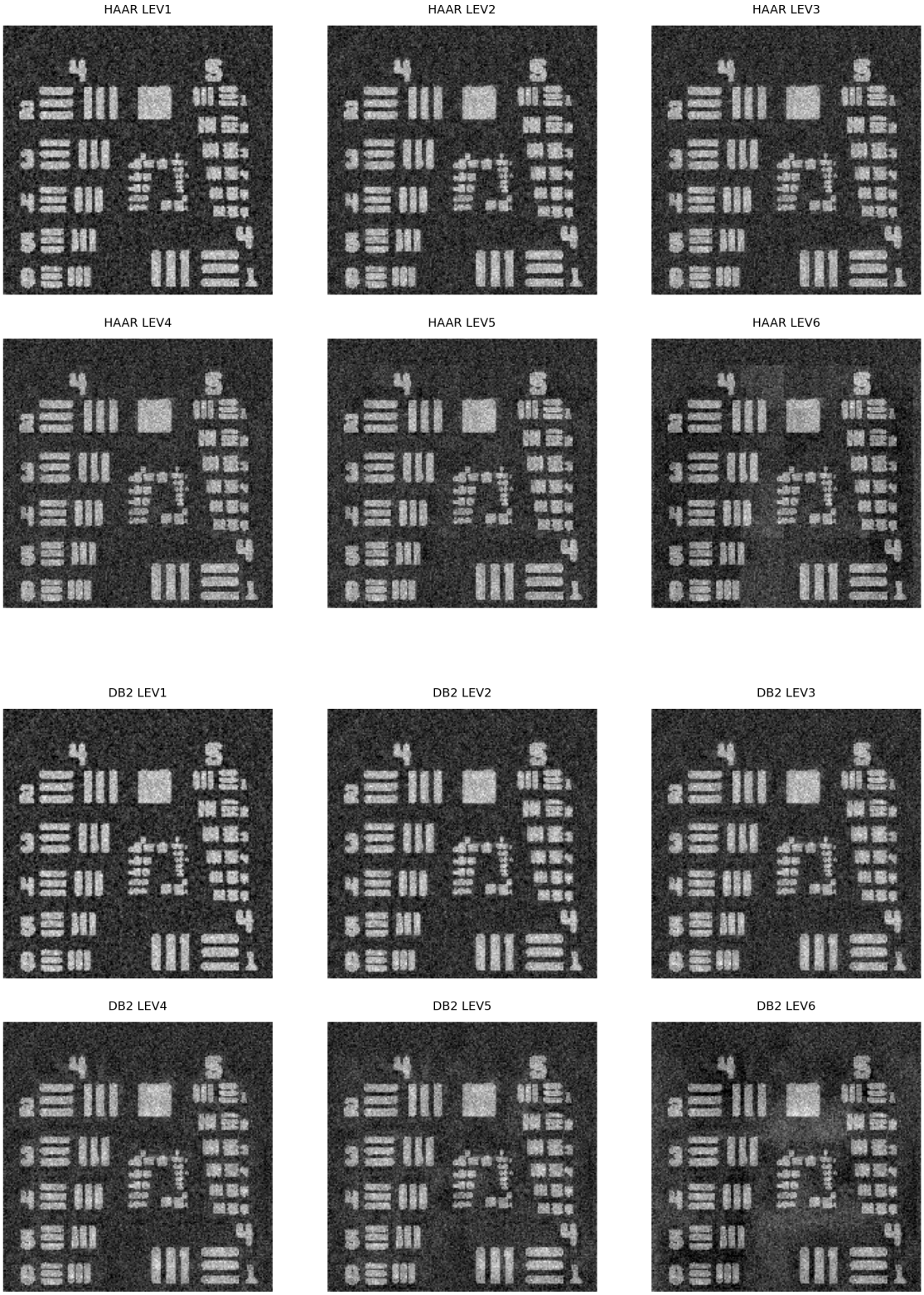


Figure 26: Peak Intensity Maps of the haar and db2 Wavelet for Different Decomposition Levels.

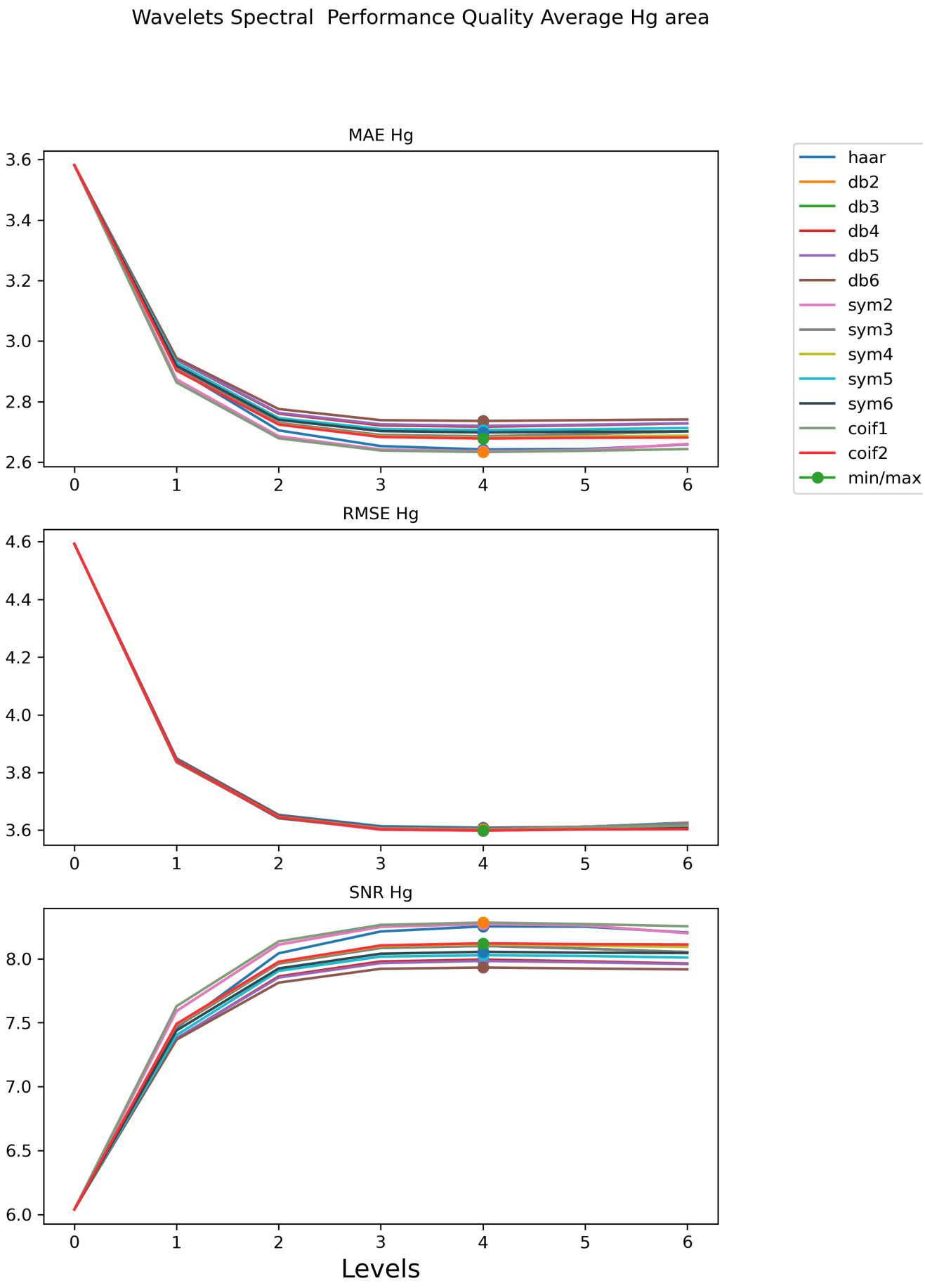


Figure 27: Average Spectral Performance Quality of Mercury L3 Peak for various wavelets against decomposition level.

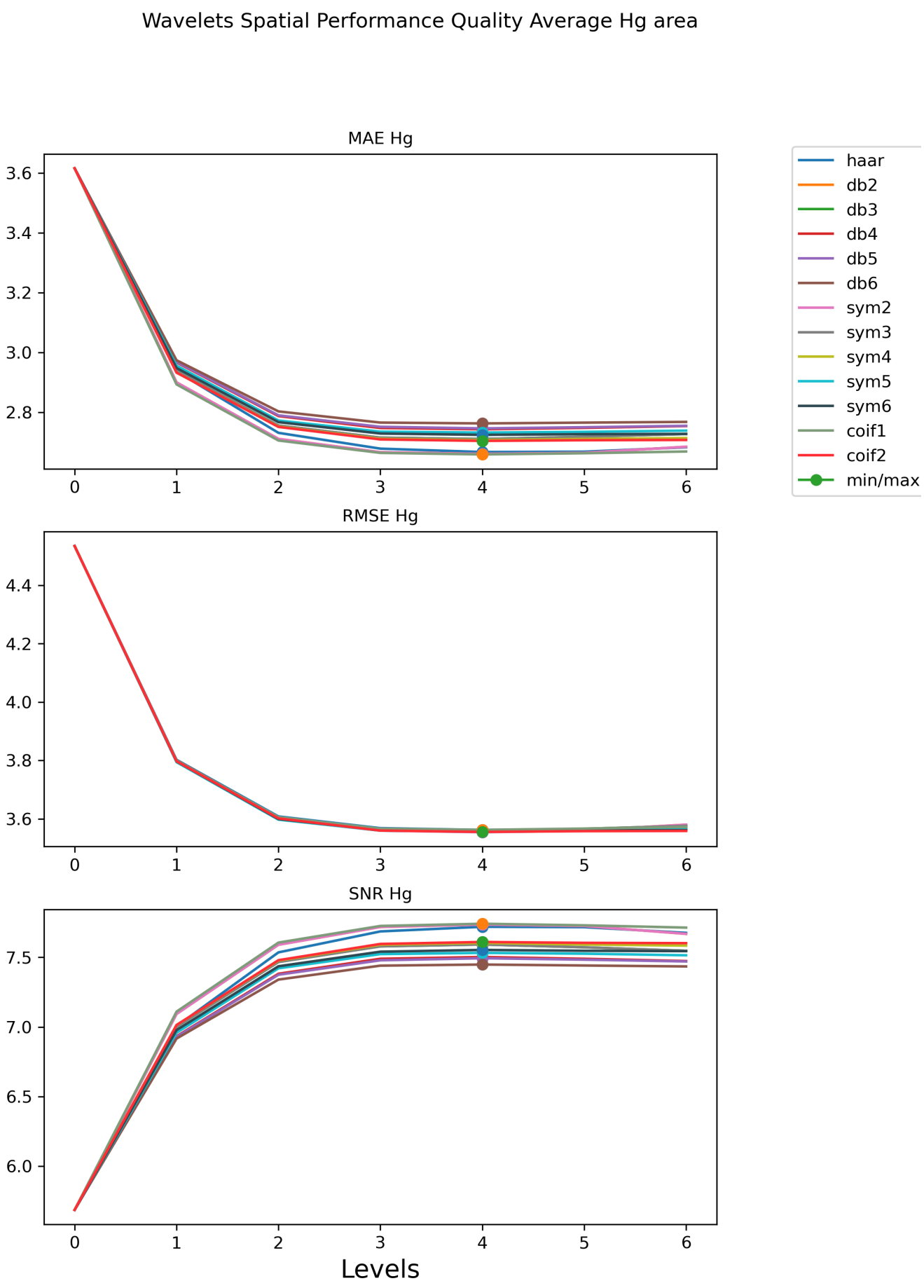


Figure 28: Average Spatial Performance Quality Mercury L3 Peak for various wavelets against decomposition level.

5.1.2 Discussion Testing Round 1

From Figure 25 it could be seen that in most cases the sym2, db2, coif1 and haar wavelets performed best. A common denominator is the small amount of vanishing moments. The haar wavelet only has the zeroth vanishing moment, while coif1, sym2 and db2 all have 2 vanishing moments. Although the support size also lowers with the amount of vanishing moments, the support size does not distinguish the above wavelets necessarily, since coif1 has a support of 6 and, therefore, has the same support size as the db3 and sym3 wavelets. What could be concluded is that not the density of singularities within the signal, but the irregularity of the signal might be decisive for the optimal wavelet choice. In section 3.3.1 we saw that a higher number of vanishing moments is preferred to create more small-detail coefficients, however, this is only optimal for regular signals. Since the MA-XRF signal is not regular, it is better captured with irregular wavelets, having a low number of vanishing moments.

This is at least the case for the spectral performance quality and the spatial average performance quality. For the spatial sum performance quality, the db2, sym2 and coif1 perform the worst. This is not surprising since the summation of the channels results in a generally more regular signal, where the irregularities that can be captured with the wavelets having low vanishing points become superfluous and wavelets with more vanishing moments and larger support appear to be optimal. Contrarily, the Haar wavelet does perform really well in the case of spatial average performance quality. A very plausible reason for this is the binary nature of the synthetic dataset, having sharp transitions, which is easy to capture with the haar wavelet.

When visualising the filtered data by mapping the Hg L3 peak intensity distribution of the dataset, artefacting could be seen when the signal was decomposed and thresholded across 6 levels. Indeed, when plotting the spectral average performance quality of the mercury peak, it became clear that the performance quality actually decreased after 4 decomposition levels. This discrepancy with Figure 27 is logically explained by the fact that the synthetic dataset is created from a very homogeneous dataset. In other words, in most channels, there is not a clear distinguishable pattern in photon intensity. When decomposing the signal into approximation and detail coefficients, in the first level there is still noise in the approximation coefficients present, while at the sixth level, part of the signal is present in the detail coefficients. When the channel is smooth, the approximation coefficients hold practically all the signal, because there is no deviation from the approximation, hence, an increased decomposition level only makes the channel signal more smoothed out and the performance quality improves. Nevertheless, in the case that there is a pattern within the channel, an extra decomposition level could cause important details to be filtered out. Hence, since Figure 25 con-

tains the performance quality of all the channels, of which a lot are smooth, the higher decomposition level seemed to perform better. While, actually, in order to distinguish important features of the dataset, four levels appear to be optimal.

Recall that the objective of this research is to improve the spectral signal-to-noise ratio. Therefore, the mother wavelets are chosen based on performing optimally for spectral SNR. The best-performing mother wavelets in this regard are the haar, db2, sym2 and coif1 wavelet. Furthermore, for the purpose of identifying important features within the dataset, a four-level decomposition seems optimal.

Based on our findings, 9 wavelet filters are chosen to determine the optimal amount of neighbouring channels in the second testing round. Since the mother wavelets haar, db2, sym2, and coif1 performed well, it is logical to choose out of them. since sym2 and db2 are similar, only db2, haar and coif1 will be tested. These will, furthermore, not only be tested for decomposition level 4 but for decomposition level 2 and 6 as well in order to test ideal s for different quality wavelet filters. This comes down to filters F.2, F.4, F.6, F.8, F.10, F.12, F.68, F.70 and F.72, in Table 3.

5.2 Testing Round 2

In the second testing round, a set of 9 different wavelet filters is evaluated for 11 different values of s, ranging from 0 to 10. Hence, 99 filters are tested, shown in Figure 29.

5.2.1 Observations Testing Round 2

Explaining and Observing the Performance Quality Graph. Similar to the first testing round, the performance quality is based on the entire dataset. The performance quality is measured in the same way as in the first round as well. Yet, instead of plotting the performance quality against the number of decomposition levels for various mother wavelets and a fixed number of neighbouring channels $s = 5$, now, the performance quality is plotted against s for various mother wavelet/level-decomposition combinations. The performance quality at $x = 0$ represents the quality of the filtered data when no neighbouring channels are used to determine the standard deviation.

A clear trend is visible in Figure 29, the more neighbouring channels are used to determine the standard deviation, the worse the performance quality, except for the spectral sum performance quality, where it is the other way around. What also can be seen is that the haar and db2 wavelet at level 2 perform worse than the haar and db2 at level 6, following the same trend as in Figure 25.

Visualising Spatial Performance. Again, similar to the first testing round, peak intensity distribution maps of Hg L3 are plotted to verify the spatial performance. Remarkably, the trend in Figure 30,Figure 31, and Figure 32 seems to be directly opposite to that of Figure 29. Significant artifacting is visible for low values of s.

Plotting the Average Spectral and Spatial Performance Quality of the Mercury L3 Peak. To verify why the visual spatial performance seems worse with low values of s in Figure 30, Figure 31, and Figure 32 while this was not the case when regarding Figure 29, the average spectral and spatial performance quality of the mercury peak is investigated. This is done similarly to the first testing round, determining the performance quality of an area of 50 channels covering the mercury L3 peak, yet, now the performance quality is plotted against s. The spectral performance quality is visualised in Figure 33 and the spatial performance quality is visualised in Figure 34.

It can be seen that the average spectral and spatial performance of the mercury area is in accordance with Figure 29 and contradicts the visual spatial performance of Figure 30, Figure 31, and Figure 32.

Plotting Detail and Approximation Coefficients. Since the numerical values of the average performance quality contradict the visual analysis, it could be the case that the filter is filtering out signal values as well as noise, over-smoothing the data. Therefore, it is interesting to regard the approximation and detail coefficients for different values of s. The approximation and horizontal detail coefficients at channel 1080 (center of the mercury L3 peak) of the row indicated by the red line in Figure 36 of the db2 lev4 wavelet filter are plotted in Figure 35.

It can be seen that indeed, at $s = 0$ the detail coefficients have been thresholded out. While, starting around $s = 4$, the detail coefficients represent the intensity changes of the red line in Figure 36 better.

Plotting the Summed Spectral and Spatial Performance Quality of the Mercury L3 Peak. Since determining the performance quality per pixel or channel within the mercury area is not in accordance with the visual representation, the spatial and spectral performance quality of the summated Hg L3 area is determined. The spectral performance quality of the summated Hg area is visualised in Figure 37 and the spatial performance quality of the summated Hg area is visualised in Figure 38. These graphs seem to be in accordance with the visual spatial performance in Figure 30, Figure 31, and Figure 32.

It can be seen that the spectral performance stays the same or increases with more neighbouring channels. The spatial performance quality is more hyperbolical

and initially strongly increases after it slightly decreases with increasing s. In both spectral and spatial performance quality, it can be seen that the higher the decomposition level, the higher the optimal value of s.

Explaining and Observing the Performance Quality Sheet for Varying s. In Figure 39, Figure 40, and Figure 41, the performance quality of filters with a db2 wavelet and a 4-level decomposition, for three different values of s are plotted. Every sheet has a similar layout. In the top row, from left to right the ground truth mercury L3 peak intensity map is shown (as a reference), next to that is the peak intensity map of the filtered data. At the top right corner, there is a small table where the performance quality of the summed data is presented. Here the spectral sum performance quality is determined by first summing all the spectra of the filtered data and then comparing that to the sum spectrum of the ground truth data. Similarly, the spatial sum performance quality is determined by first summing all the channels of the filtered data and then, comparing that to the sum of channels of the ground truth data.

In the leftmost column, the spectral performance quality is plotted for every pixel. Similarly, the spatial performance is plotted for every channel in the middle column, giving an overview of the spectral performance quality per pixel and the spatial performance quality per channel, respectively. In the right column, the spectral performance for the mercury L3 peak is plotted, based on 50 channels covering the peak.

Above all nine performance quality plots, the average performance is indicated (AVG). For the spectral performance quality plots also the standard deviation is indicated (SIG).

When regarding the numerical values in the figures, it can be seen that the spectral performance quality for the summed data remains constant for different s. Meanwhile, the spatial performance quality for the summated data and the spectral and spatial performance for the average data get worse with increasing s. Hence, the numerical values are in agreement with Figure 29, Figure 33, and Figure 34. Yet, when regarding the spectral performance for the mercury L3 peak, in the rightmost column, a possible explanation is given as to why the numerical error values are not in agreement with Figure 30, Figure 31, and Figure 32. In Figure 39, where $s = 0$, the performance quality is very high in the homogeneous areas of the mercury distribution pattern, yet, it is significantly worse at its edges. The standard deviation of the error values is also significantly higher for $s = 0$ than for $s = 5$ and $s = 10$.

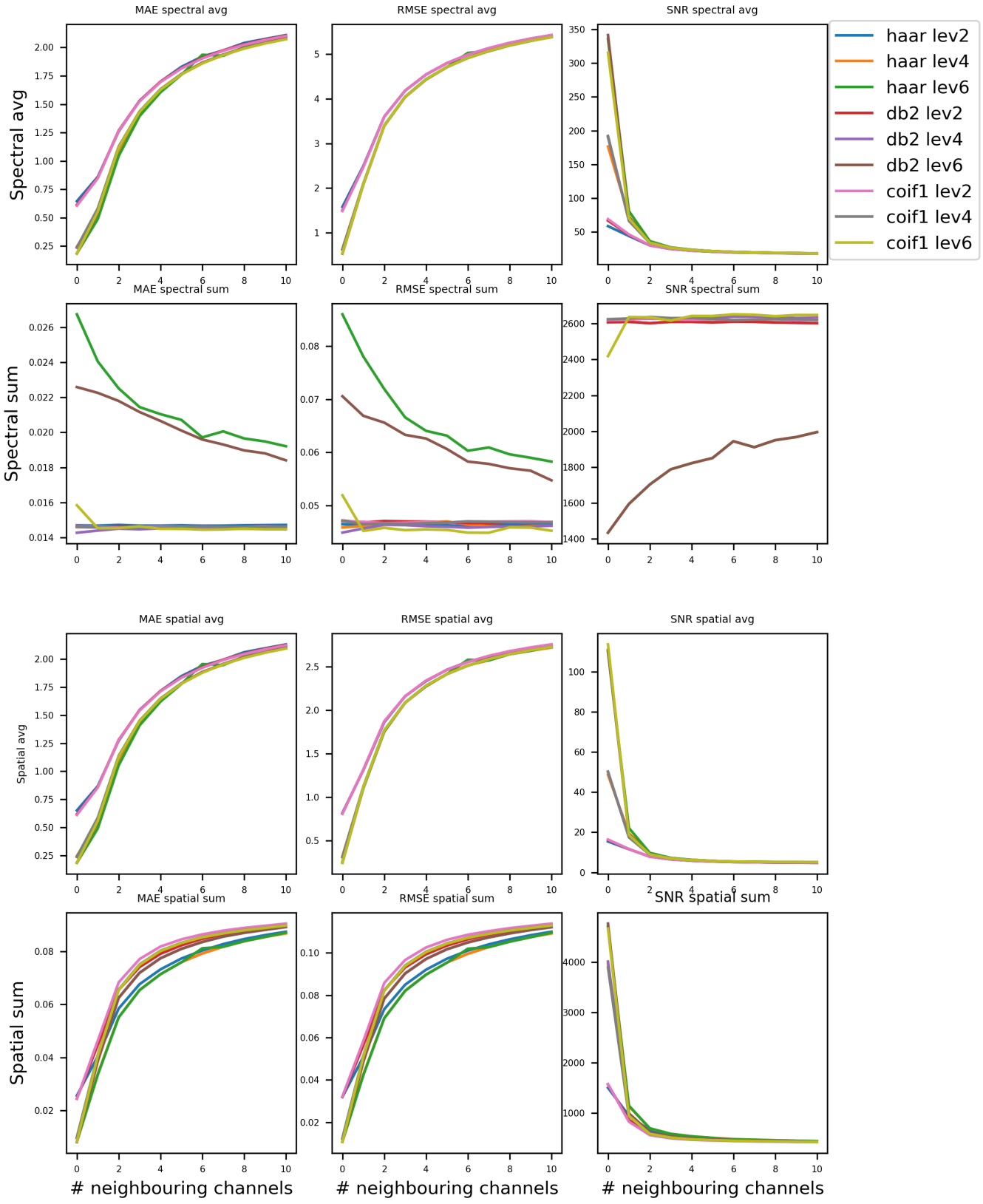


Figure 29: Filter performance, indicated by MAE, RMSE, and SNR for the haar, db2, and coif1 wavelet at decomposition level 2,4, and 6 for increasing number of neighbouring channels.

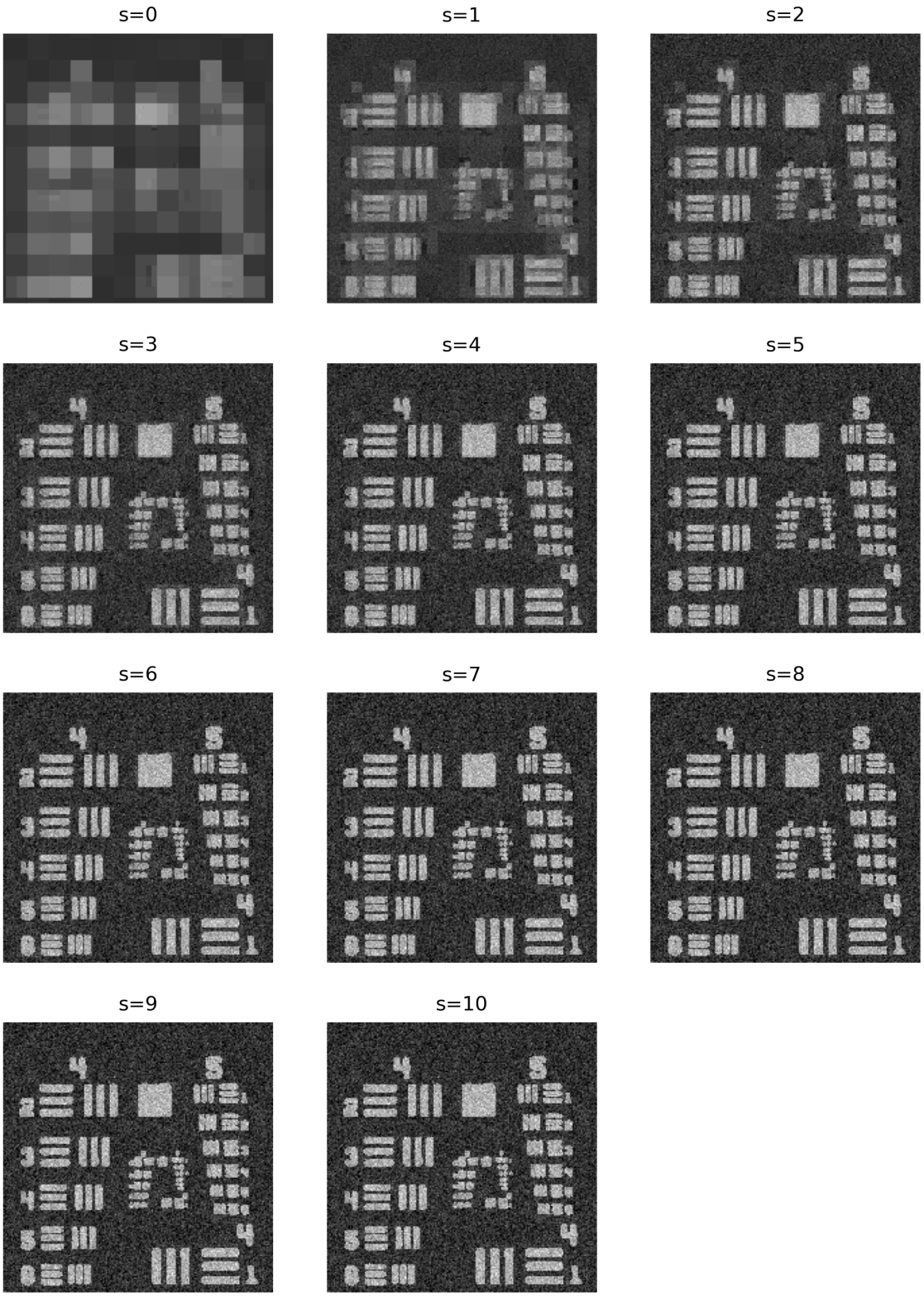


Figure 30: Peak Intensity Maps of the haar wavelet at level 4 for increasing number of neighbouring channels

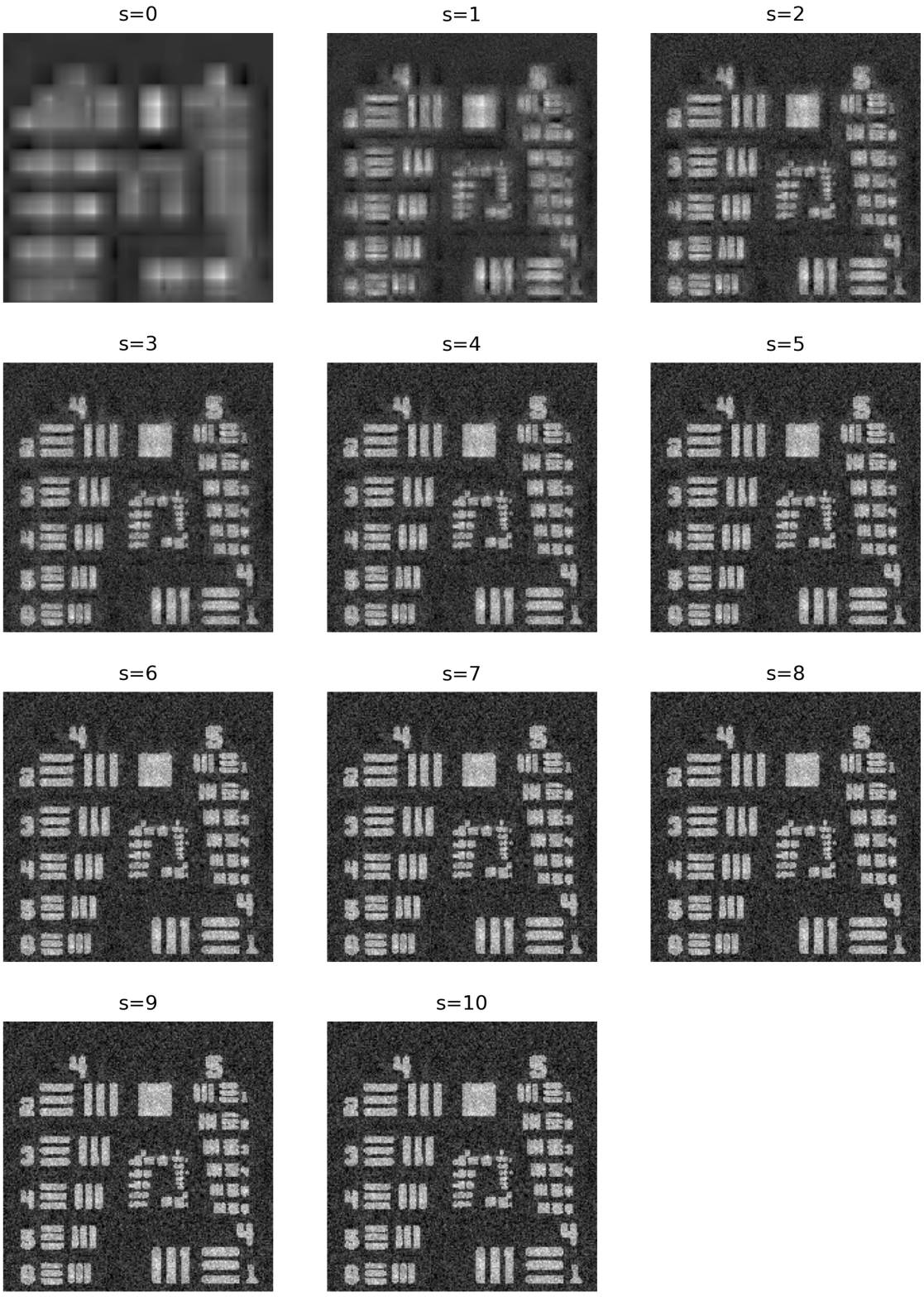


Figure 31: Peak Intensity Maps of the db2 wavelet at level 4 for increasing number of neighbouring channels

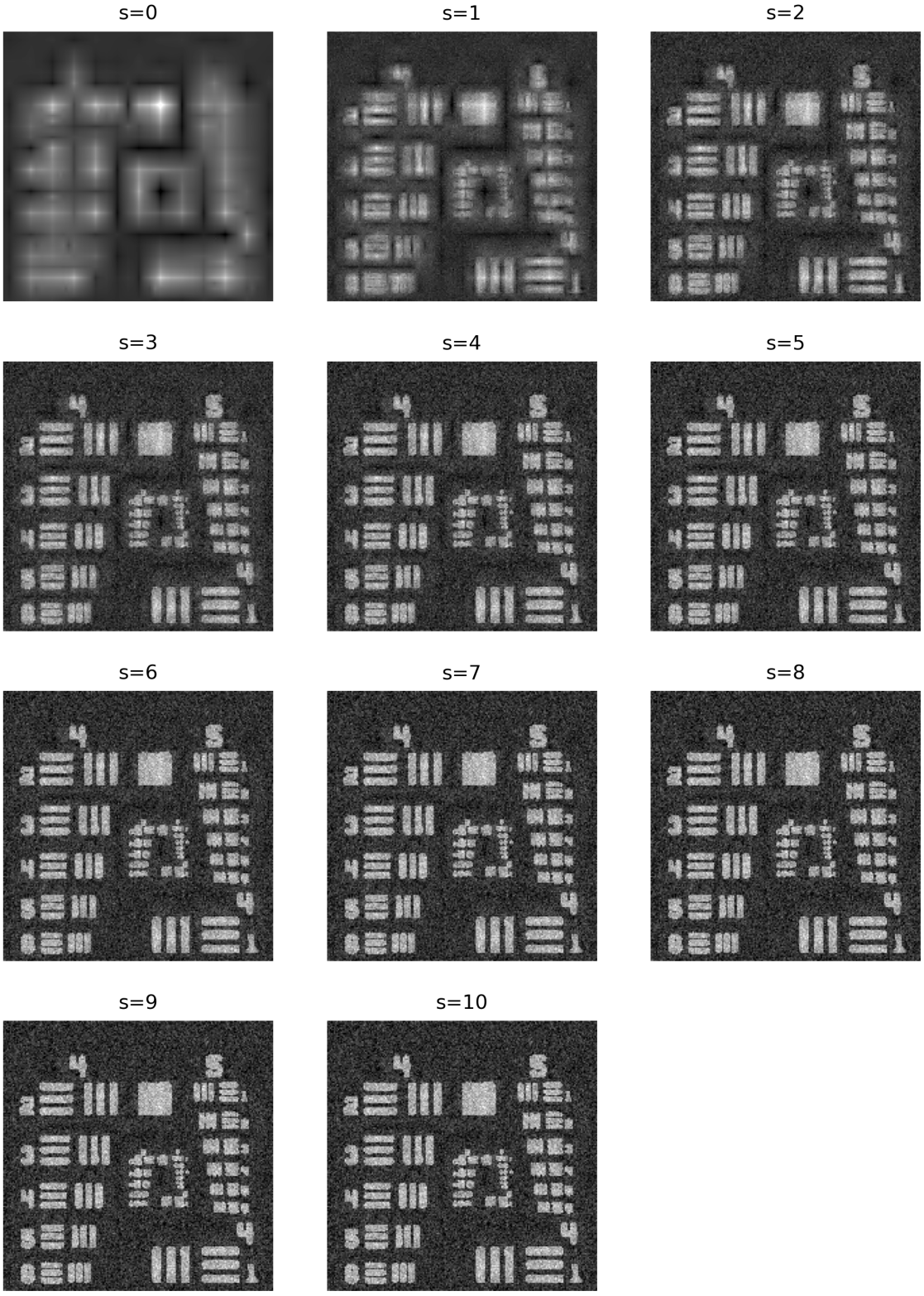


Figure 32: Peak Intensity Maps of the coif1 wavelet at level 4 for increasing number of neighbouring channels

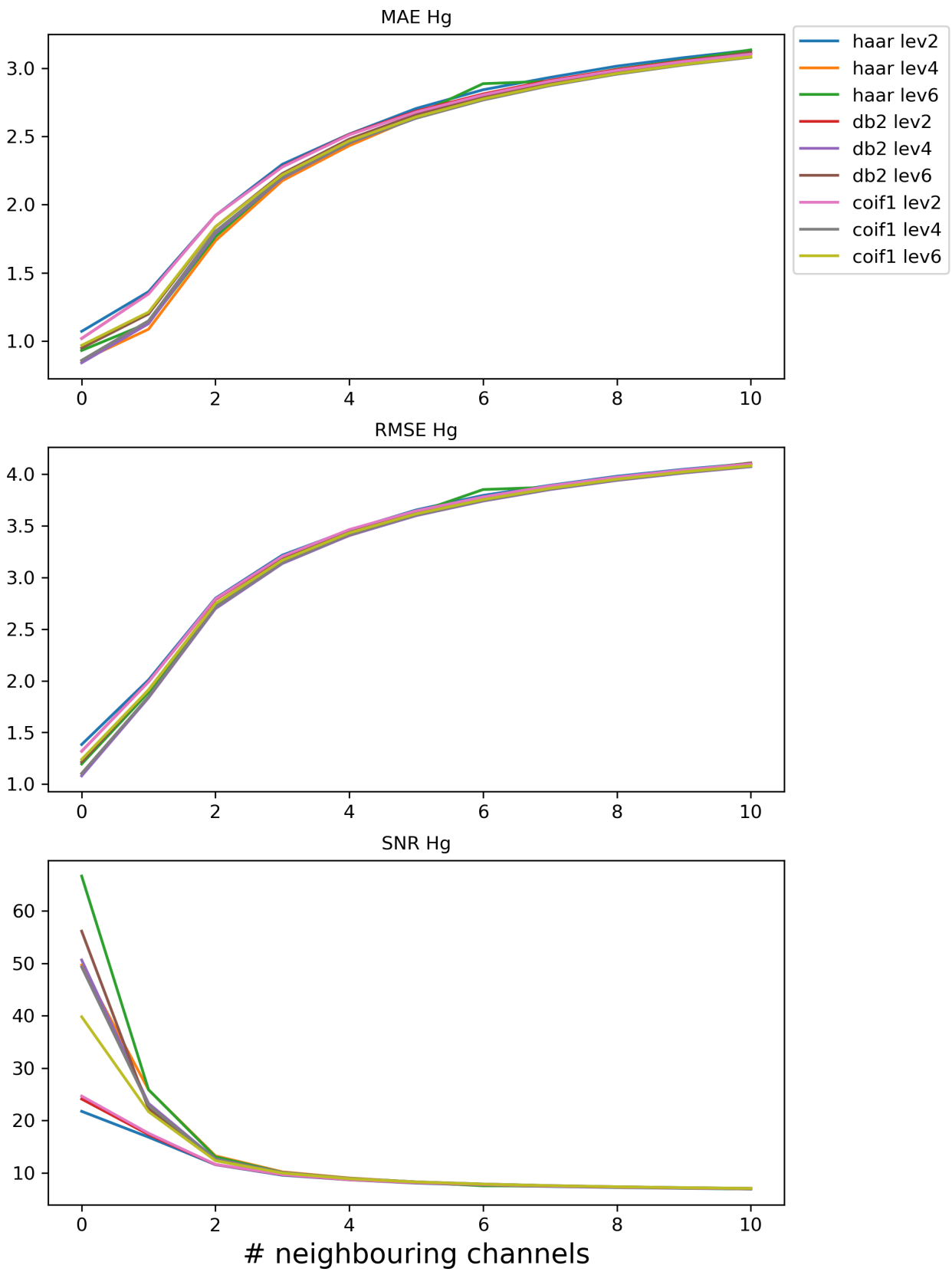


Figure 33: Average Spectral Performance Quality of the Hg L3 Area For different decomposition levels plotted against s

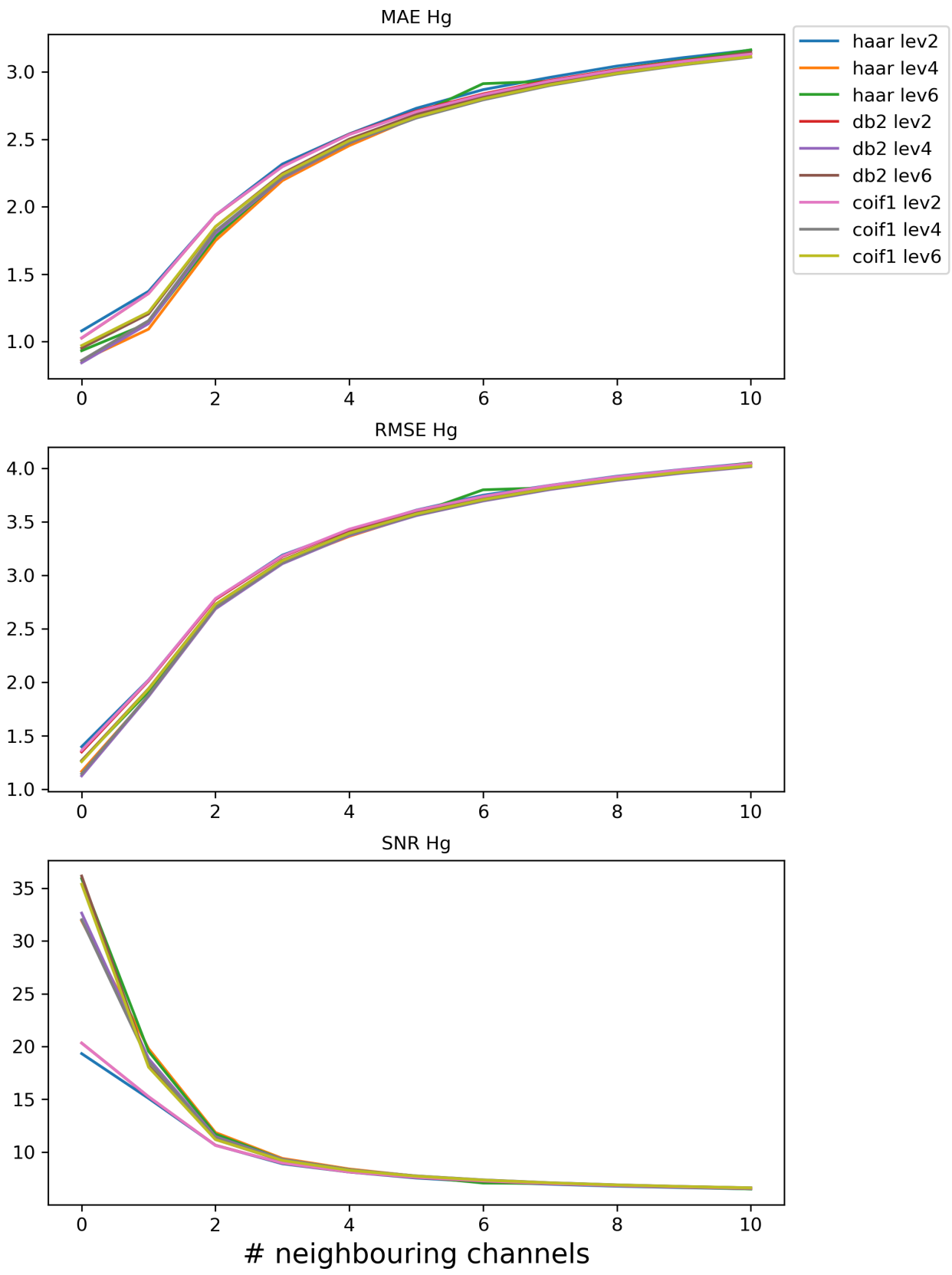


Figure 34: Average Spatial Performance Quality of the Hg L3 Area For different decomposition levels plotted against s

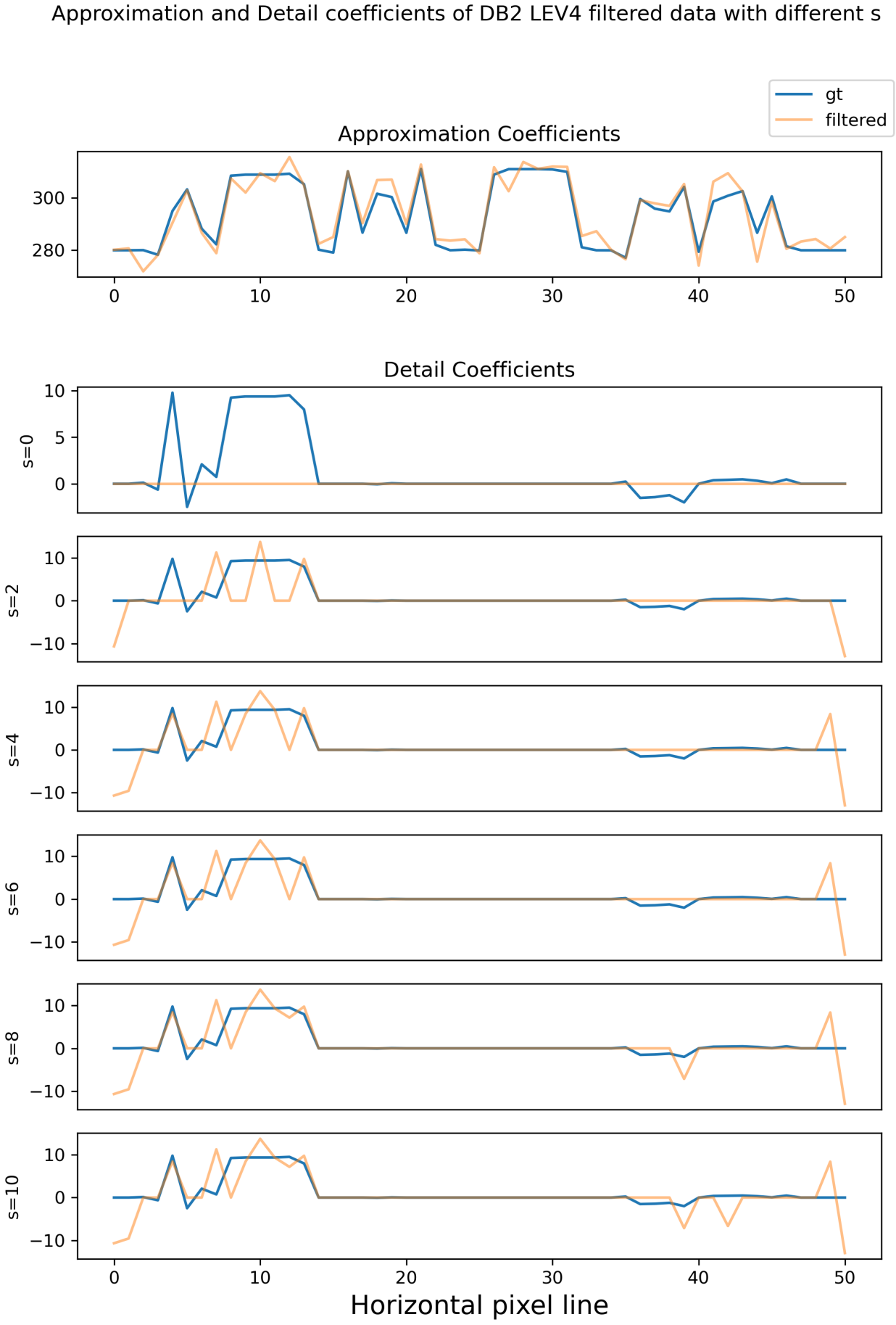


Figure 35: Approximation and Horizontal Detail Coefficients of a pixel row (see Figure 36 at channel 1080 obtained with db2 at level 4 for various s

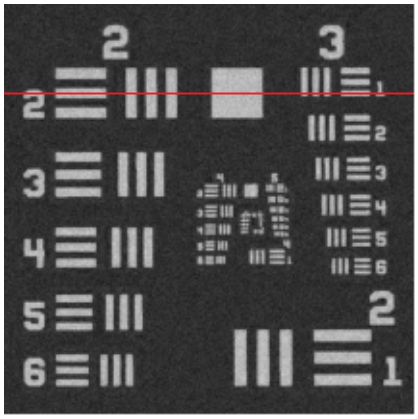


Figure 36: Row in red from which the detail and approximation coefficients are visualised in *Figure 35*.

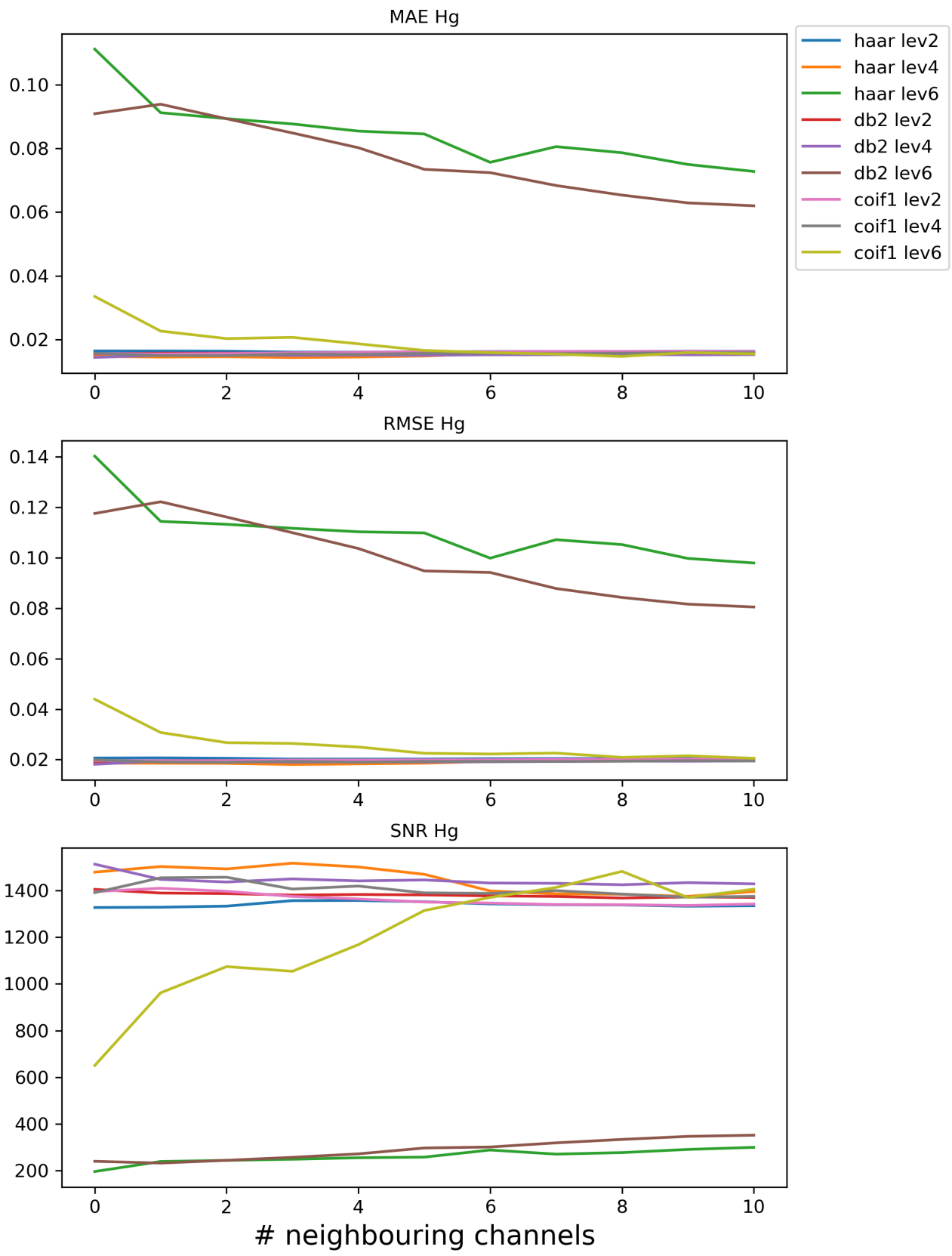


Figure 37: Summated Spectral Performance Quality of the Hg L3 Area For different decomposition levels plotted against s

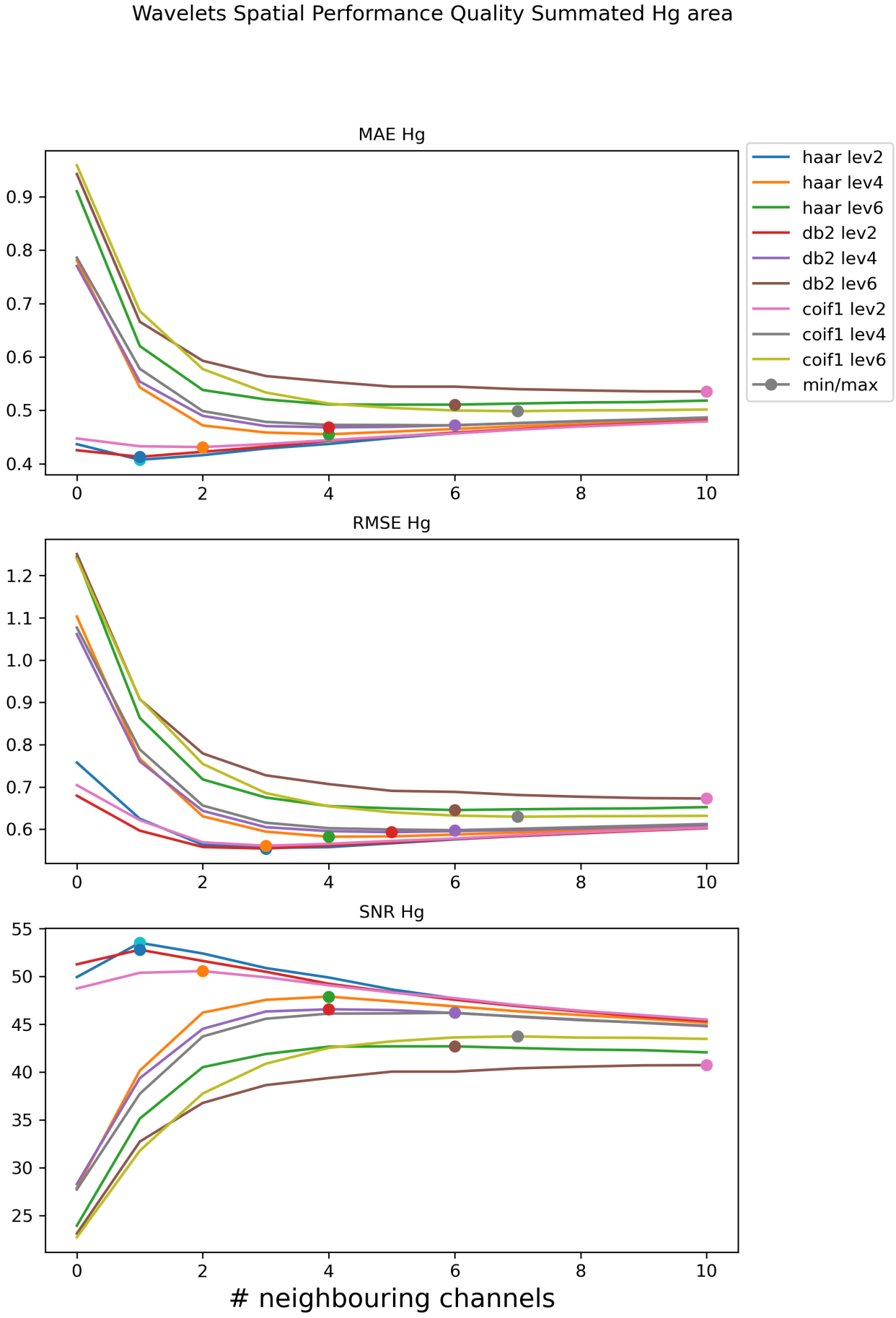


Figure 38: Summated Spatial Performance Quality of the Hg L3 Area For different decomposition levels plotted against s

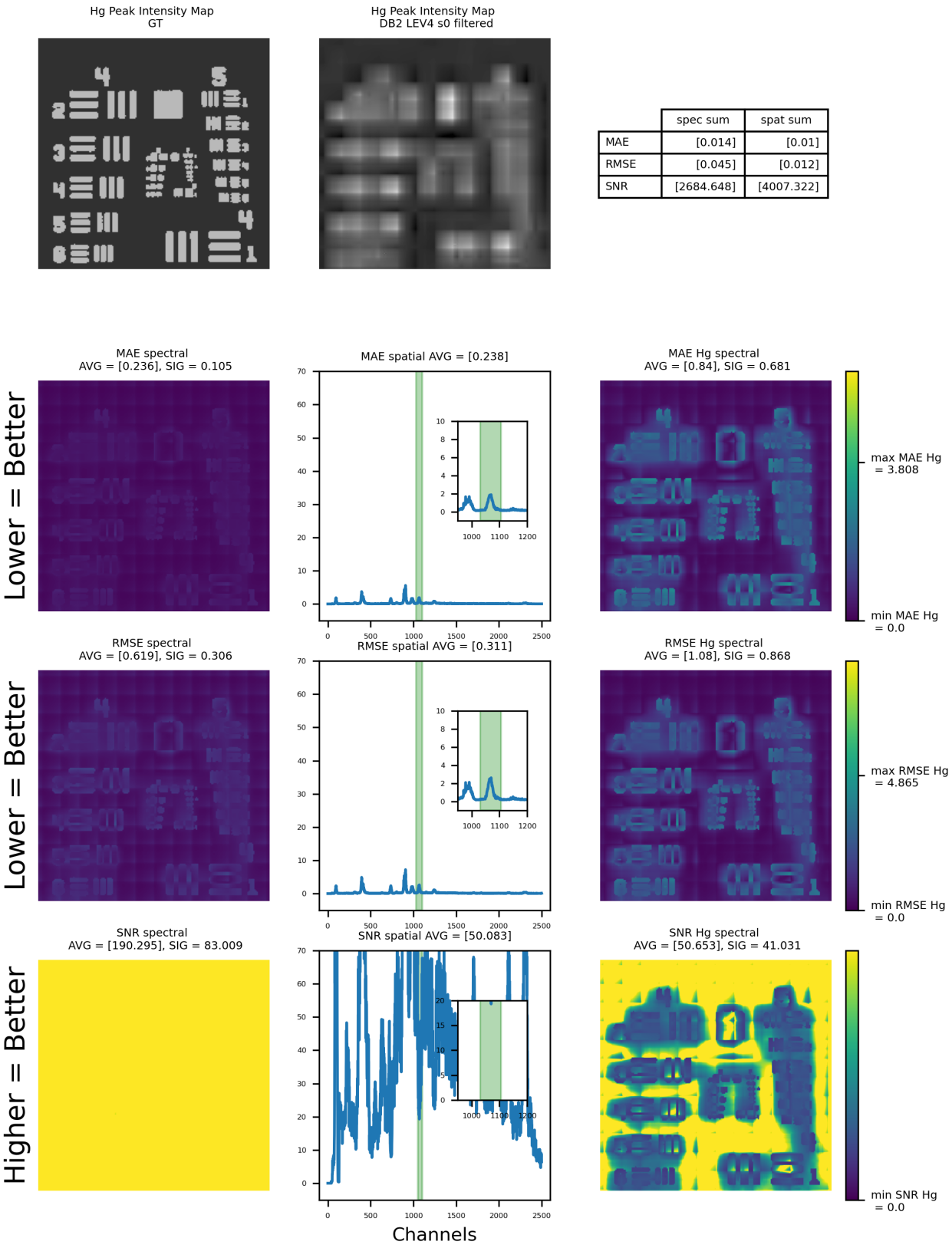


Figure 39: Performance Quality of filter applying a db2 wavelet at decomposition level 4 with $s = 0$

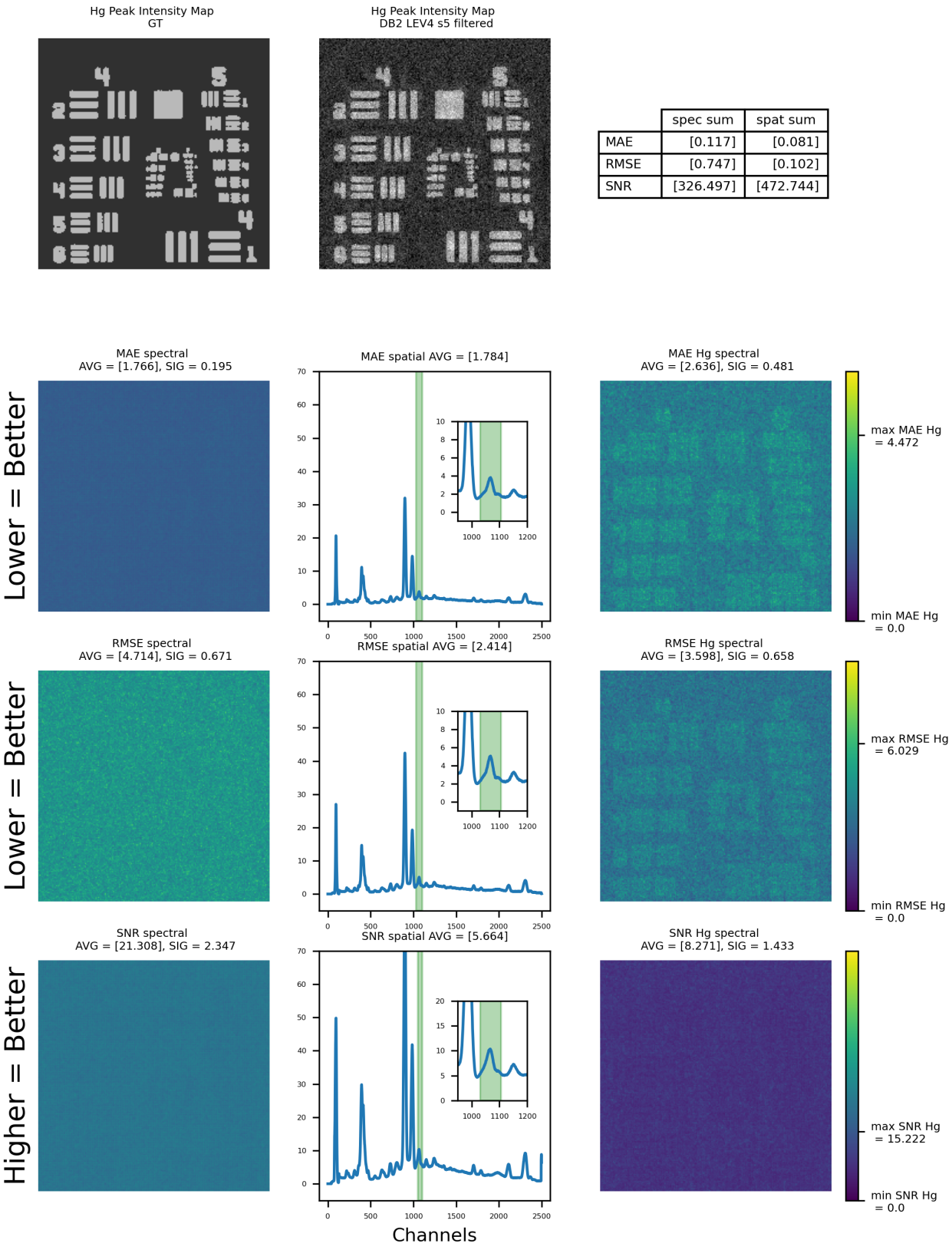


Figure 40: Performance Quality of filter applying a db2 wavelet at decomposition level 4 with $s = 5$

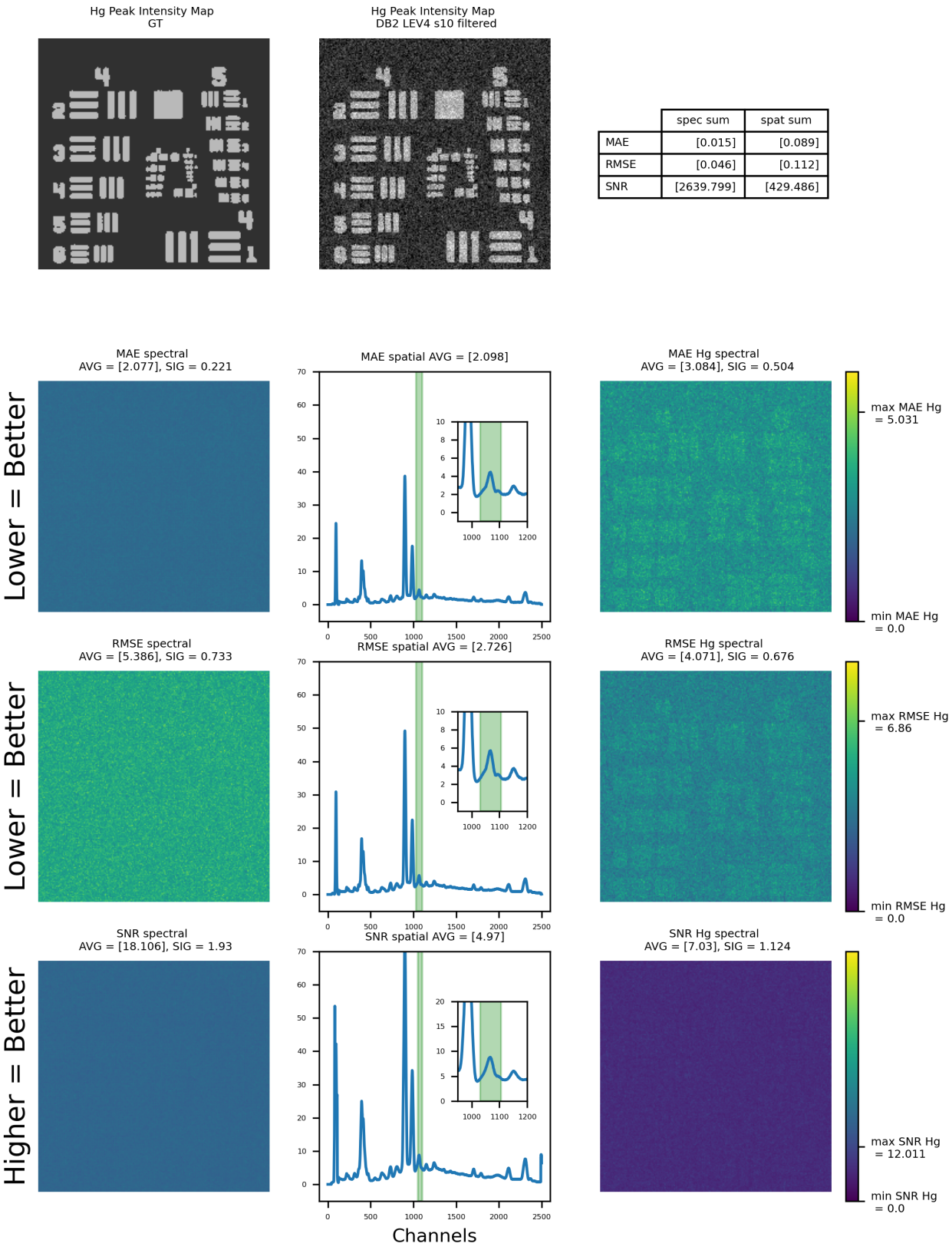


Figure 41: Performance Quality of filter applying a db2 wavelet at decomposition level 4 with $s = 10$

5.2.2 Discussion Testing Round 2

A clear trend stood out regarding the numerical performance quality indicators. Figure 29, Figure 33, and Figure 34: all seemed to indicate better performance at lower values of s. However, a visual inspection of the mercury L3 peak intensity maps from Figure 30, Figure 31, and Figure 32 indicated exactly the opposite.

A probable explanation can be found in the nature of the MA-XRF data and the height of the threshold values. The Poisson distribution of the data causes large intensity fluctuations. When the standard deviation of a channel is determined only on the data of that single channel, which has a small amount of counts, the standard deviation will be large. A large standard deviation results in a large threshold and, therefore, also detail coefficients with higher values will be set to zero. Figure 35 shows that indeed, for lower values of s, more detail coefficients are set to zero. When removing all the detail coefficients, we are basically averaging the data. This gives good numerical performance quality, however, often results in failing to capture important singularities within the signal.

Starting with Figure 29, it can indeed be seen that both the average spectral and spatial performance quality decrease with increasing s. The effect of shot noise is, however, less prominent and actual intensity differences from the signal within the data become more prominent when first summing the data spatially or spectrally before calculating the performance quality. Accordingly, with higher s the spectral sum performance quality decreases. The reason why this is not the case for the spatial sum performance quality probably has to do with the fact that outside the mercury distribution peak, most elemental distributions are largely homogeneous. Hence, over-averaging does not have a negative influence on the spatial performance quality indicators.

Just like the first testing round, we zoomed in on the mercury peak to enlarge the effect the filter has on areas with a higher amount of singularities. Where in the first testing round a clear difference was obtained between the average performance quality of the entire dataset and the average performance quality of the mercury L3 area, this was not the case in the second testing round. Figure 33 and Figure 34 show similar behaviour as the first and third row of Figure 29 respectively. While both a high decomposition level and low s cause averaging to occur, the nature of the averaging is different, which explains why the performance quality of Figure 27 and Figure 28 drops at a certain point when too much is averaged, while this is not the case for low s as seen in Figure 33 and Figure 34: At higher level decomposition, the scaling and wavelet functions become larger, hence, the approximation coefficients and detail coefficients represent larger areas. Consequently, in the process of reconstruction, not only small deviations are filtered out, but also part of the signal gets redistributed. The latter can cause the performance quality to drop. When s is low, in other words, the threshold is

high and the detail coefficients are almost all set to zero, the only thing left is the approximation coefficients. Hence, only local averaging occurs and we see that the redistribution depends on the decomposition level alone.

Still, a low s does result in artefacts as was clear from Figure 30, Figure 31, and Figure 32. In order to capture this numerically, the summated spectral and spatial performance quality of the Hg area was determined and visualised in Figure 37 and Figure 38. The spectral performance increases or stays constant with increasing s, while the spatial performance first strongly increases after it slightly decreases with increasing s. The most likely reason for the behaviour of the latter has to do with the increasingly lower threshold with increasing s. When the threshold decreases, initially, important details of the signal are being preserved instead of discarded, however, after a certain point, the threshold becomes so low that also noise is preserved. Another observation was that, the higher the decomposition level, the higher the s for optimal performance. To this end, a testing round 2b was initiated, giving insight into the codependence of level decomposition and neighbouring channels.

The effects of different s were, moreover, clarified by plotting the error values spatially and spectrally for the db2 wavelet filter with a 4-level decomposition for varying s in Figure 39, Figure 40, and Figure 41. Here, the spectral performance quality distributions of the mercury area indicated how the average performance could be good, while the images were distorted for low s, showing large error differences at the edges of the mercury distribution pattern. It is visible that the performance quality is very high for the homogeneous areas, in accordance with the over-averaging that occurs at low s. Another observation is that there is a large visual improvement of the peak intensity maps and the rightmost column and a significant decrease in the standard deviation σ between $s = 0$ and $s = 5$. However, between $s = 5$ and $s = 10$, these improvements are not as obvious, meanwhile, the numerical performance indicators become worse. When regarding Figure 38, it can indeed be seen that the significant improvements are till $s = 3$ and then it flattens out. As previously mentioned, this probably has to do with noise not being discarded when the threshold is too low.

5.3 Testing Round 2b: On the Interaction of Decomposition Levels and Neighbouring Channels

From Figure 38 it became clear that the optimal value for s depends on the number of decomposition levels. Therefore, in order to check that our outcome of the first testing round is not biased by initially taking $s = 5$, again the spectral and spatial performance quality of the Hg area is plotted against the number of decomposition levels. Yet, instead of filters having various mother wavelets, now filters have varying s values. Moreover, peak intensity maps for different combinations of level decompositions and values s are plotted. Finally, detail and approximation coefficients are plotted for various s and decomposition levels.

5.3.1 Observations Testing Round 2b

Plotting the Average Spectral and Spatial Performance against Decomposition Level for the Hg L3 Area. The average spectral and spatial performance of the Hg L3 area was plotted against the number of decomposition levels for haar, db,2 and coif1 with various s, shown in Figure 42, Figure 43, Figure 44, Figure 45, Figure 46, and Figure 47. It can be seen that, generally, optimal performance is still achieved at level 4, the one outlier being for s=0.

Plotting the Summated Spectral and Spatial Performance against Decomposition Level for the Hg L3 Area. When plotting the performance quality against the number of neighbouring channels, more insight was provided by regarding the performance quality of the summed Hg area instead of looking at the average performance quality of individual spectra or channels. Therefore, also the summed performance quality is plotted against increasing levels for functions having various values of s. The summated spectral performance quality is visualised in Figure 48, Figure 49, and Figure 50 and the summated spatial performance quality is visualised in Figure 51, Figure 52, and Figure 53.

The consistency of level 4 being optimal, which seemed to be the case when determining the average performance quality of the Hg L3 area does not apply to the summated performance quality. The summated spectral performance quality is optimized by applying a higher-level decomposition, while the summated spatial performance quality, is optimized by applying a lower-level decomposition. Moreover, it can be seen that for coif1 the optimal summated spectral performance for low s is at a lower level decomposition and an optimal performance for high s is at a higher level decomposition, this is not the case for the haar and db2 summated spectral performance where an optimum for almost every number of neighbouring channels is obtained at a level 5 and 4 decomposition respectively.

Plotting the Peak Intensity Maps of the db2 Wavelet for Varying Combinations of Levels and Number of Neighbouring Channels of the Hg L3 Peak. From the performance quality graphs of the summated Hg areas, Figure 37, Figure 38, Figure 53, and Figure 52, it seems that the optimal number of decomposition level depends on the number s and vice versa. The peak intensity maps shown in Figure 54 indeed show that a higher number of s is preferred for a higher decomposition level. Another observation is that for $s = 0$ the visual performance is not satisfactory for any decomposition level. Similarly for a 6-level decomposition no amount of s results in a peak intensity map without artefacting. The blue line circles the peak intensity maps that do obtain visually appealing results.

Plotting Detail and Approximation Coefficients. The relation between neighbouring channels and decomposition levels can not only be visualised in peak intensity maps, but also by plotting the detail and approximation coefficients. This might give more insight into the proportion of noise in the approximation coefficients and the proportion of signal in the detail coefficients at different levels and what s suits best to remove that noise. The approximation and horizontal detail coefficients at channel 1080 (centre of the mercury L3 peak) of the row indicated by the red line in Figure 36 of the db2 wavelet filter are plotted in Figure 55. It can be seen that at level 2, still a lot of noise is present in the the approximation coefficient, while at level 4 the approximation coefficient is cleaner, but part of the signal is already present in the detail coefficients. At level 6, not much is left of the original shape of the data within the approximation coefficients. The detail coefficients are all thresholded out for every level at $s = 0$, the higher the s, the more detail coefficients are contained. Which at level 2 results in more noise being kept in, and, at level 4 and level 6 seems to result in aspects of the signal being preserved.

Zooming in

From Figure 55, it could be seen that between a level 2 and level 4 decomposition, the decomposition went from approximation coefficients including noise to detail coefficients containing signal. Therefore, it is interesting to regard a level 3 decomposition as well. Moreover, from Figure 39, Figure 40 and Figure 41 it could be seen that significant improvements were made between $s = 0$ and $= 5$, unlike between $s = 5$ and $s = 10$, where, the high-frequency aspects of the spatial signal were preserved, yet the noise was not discarded. Therefore, it is interesting to regard filter outcomes applying $s = 1$, $s = 3$ and $s = 5$. To this end, three extra graphs have been made, applying a db2 wavelet.

Plotting Detail and Approximation Coefficients: Zoomed In. The zoomed-in version of Figure 55, Figure 56 shows the detail and approximation coefficients of a level 2,3 and 4 decomposition, each combined with values of s between 0 and 5. It can be seen that for a level 3 decomposition, the ground truth (blue graph) approximation coefficients are still in accordance with the horizontal pattern in Figure 36. At the same time, less noise is contained in the filtered noisy data (orange graph) than in the case with level 2.

Plotting Summated Spatial Performance against Number of Neighbouring Channels for the Hg L3 Area: Zoomed In. Although, Figure 56 gives a clear indication of what decomposition level is suitable for wavelet denoising, a proper number of neighbouring channels cannot be instantly inferred. Therefore, the spatial performance for the summated mercury area is plotted against the number of neighbouring channels for levels 2,3 and 4, visualised in Figure 57. The minima lie between $s = 1$ and $s = 5$, with the strongest improvements between $s = 0$ and $s = 2$. For a level 3 wavelet decomposition, the choice for $s = 3$ seems suitable.

Plotting the Peak Intensity Maps of the db2 Wavelet for Varying Combinations of Levels and Number of Neighbouring Channels of the Hg L3 peak: Zoomed In. To check the observations from Figure 56 and Figure 57, the peak intensity maps of the level 2,3, and 4 with an s between 0 and 5 are visualised in Figure 58. It can be seen that the maps are largely free from spatial artefacting within the blue area.

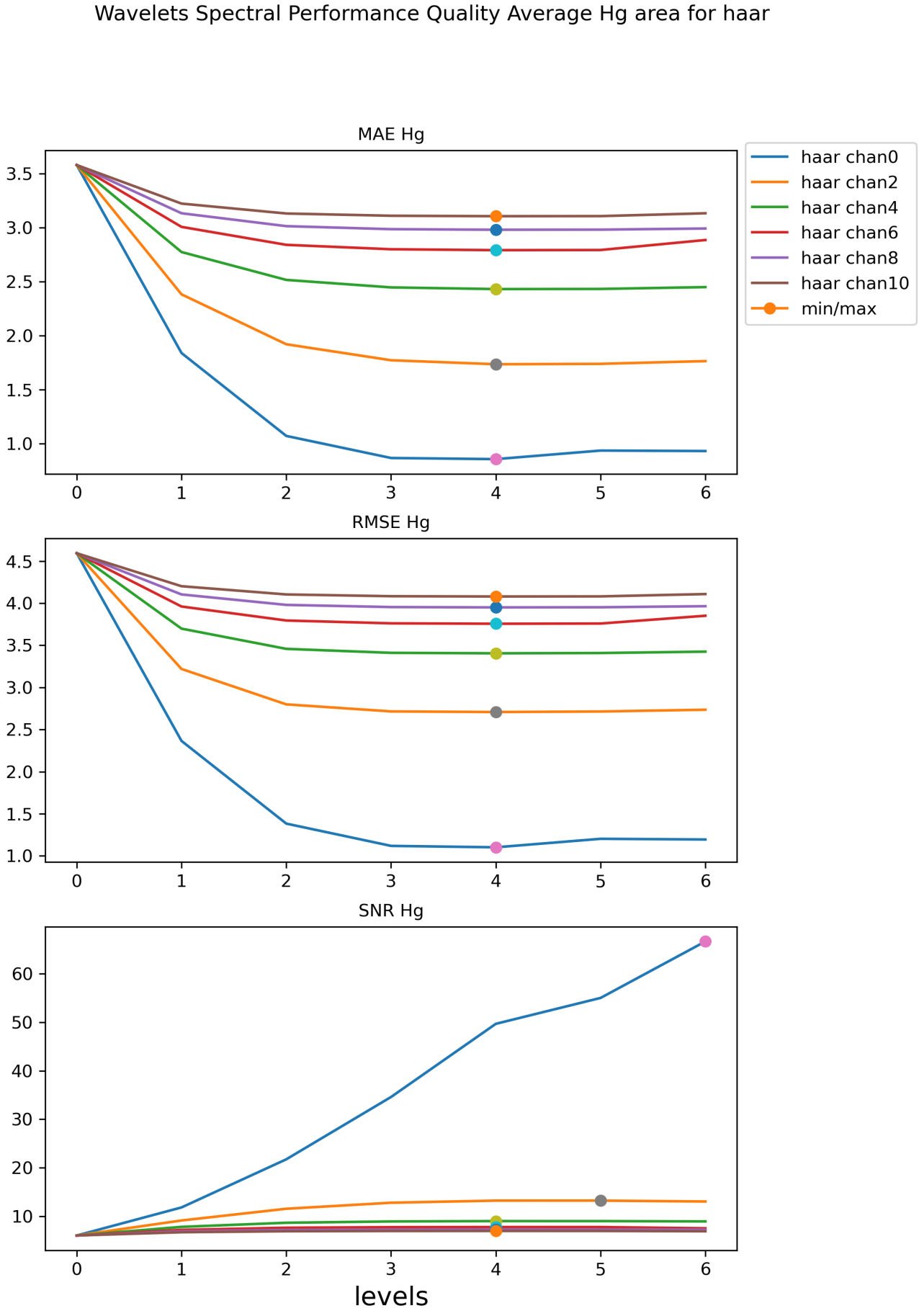


Figure 42: Average spectral performance quality of the Hg L3 Area for haar at different s plotted against the number of decomposition levels

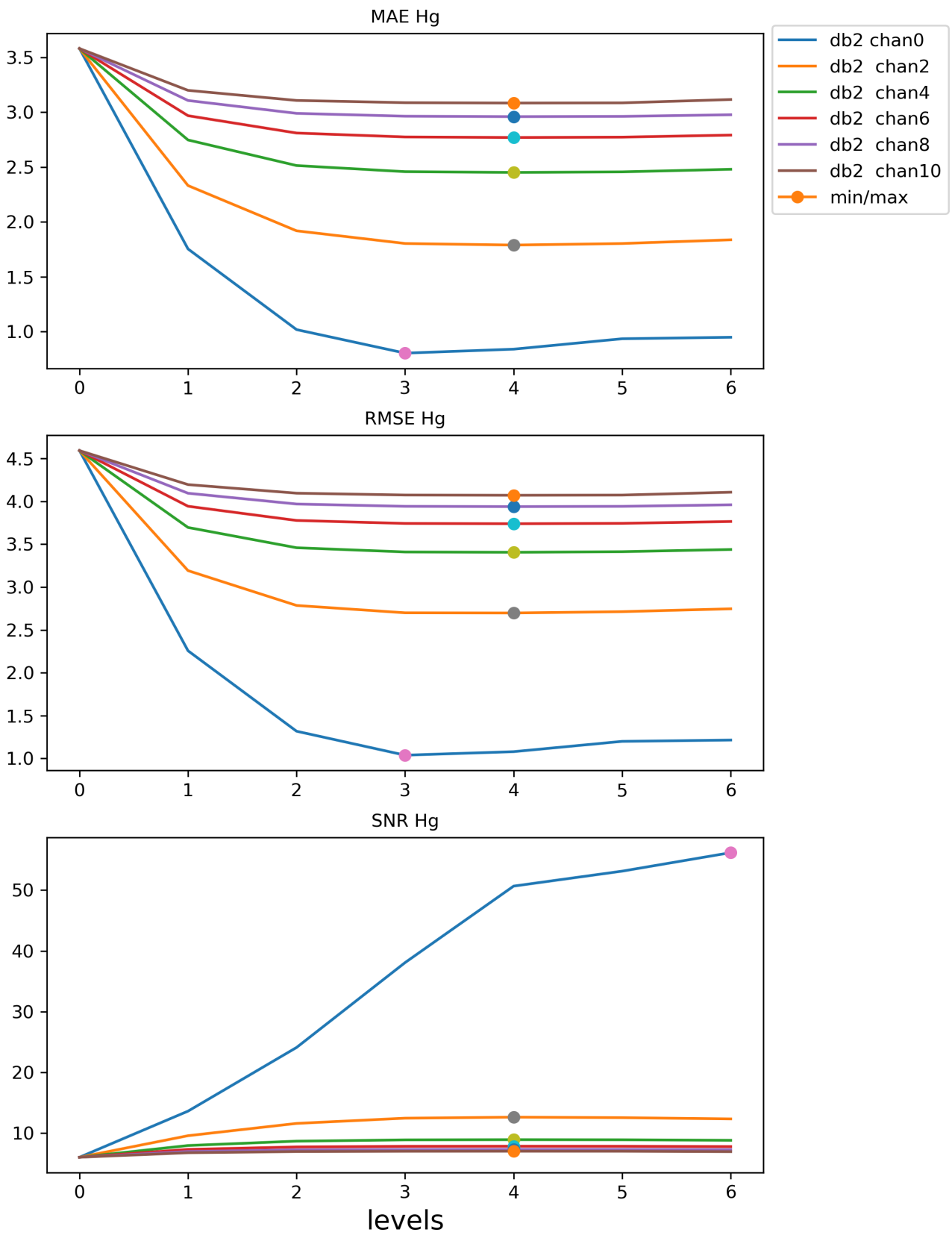


Figure 43: Average spectral performance quality of the Hg L3 Area for db2 at different s plotted against the number of decomposition levels

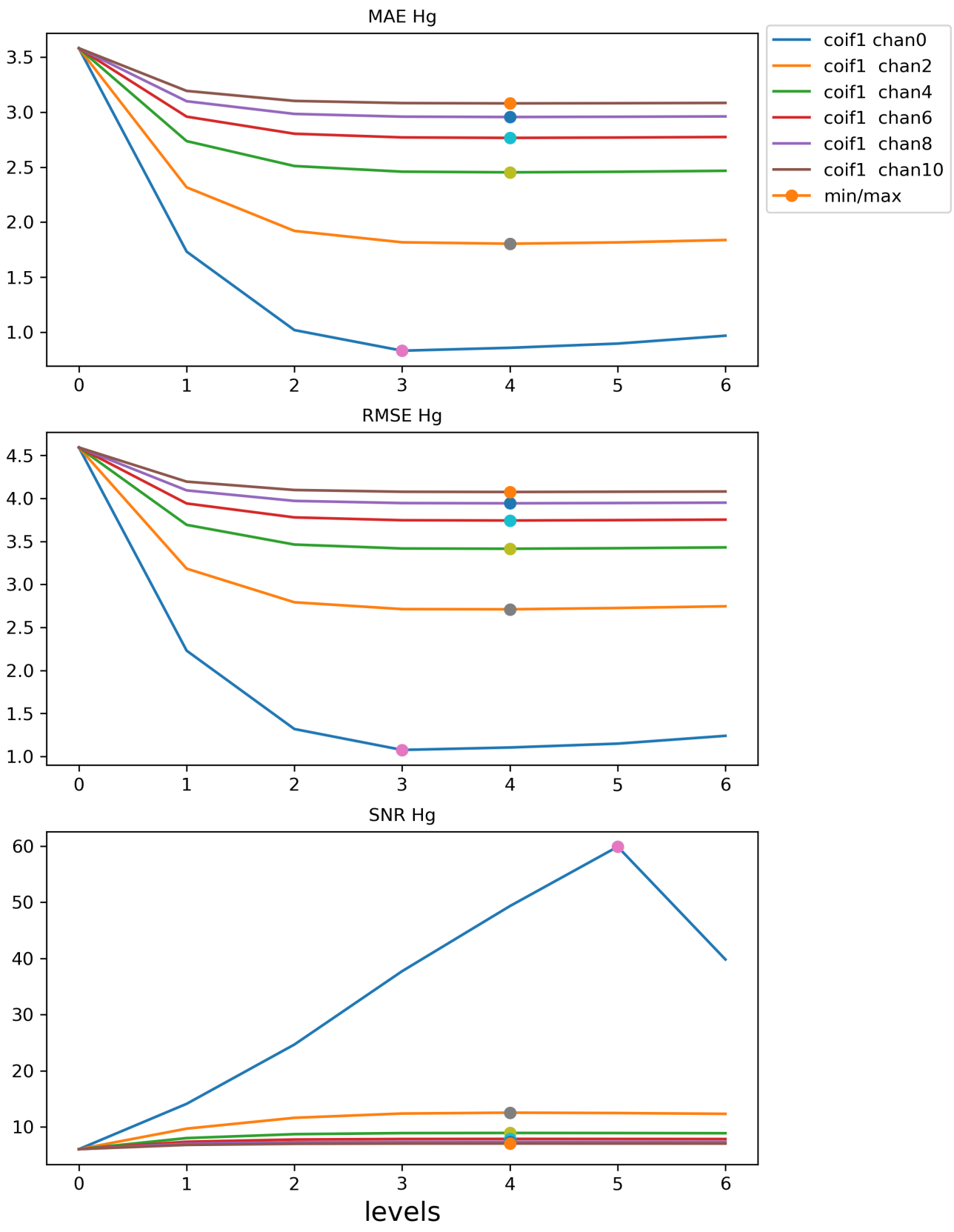


Figure 44: Average spectral performance quality of the Hg L3 Area for coif1 at different s plotted against the number of decomposition levels

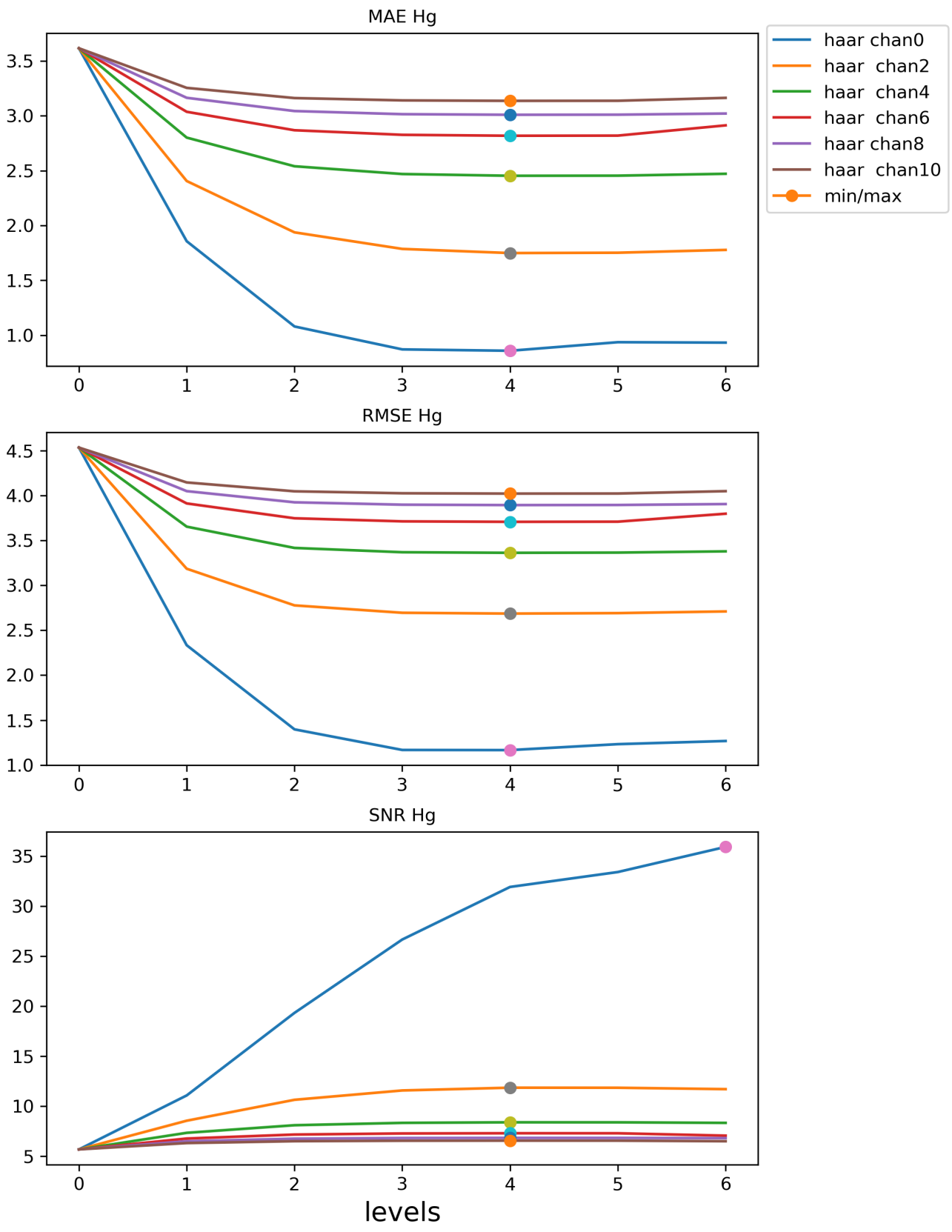


Figure 45: Average spatial performance quality of the Hg L3 Area for haar at different s plotted against the number of decomposition levels

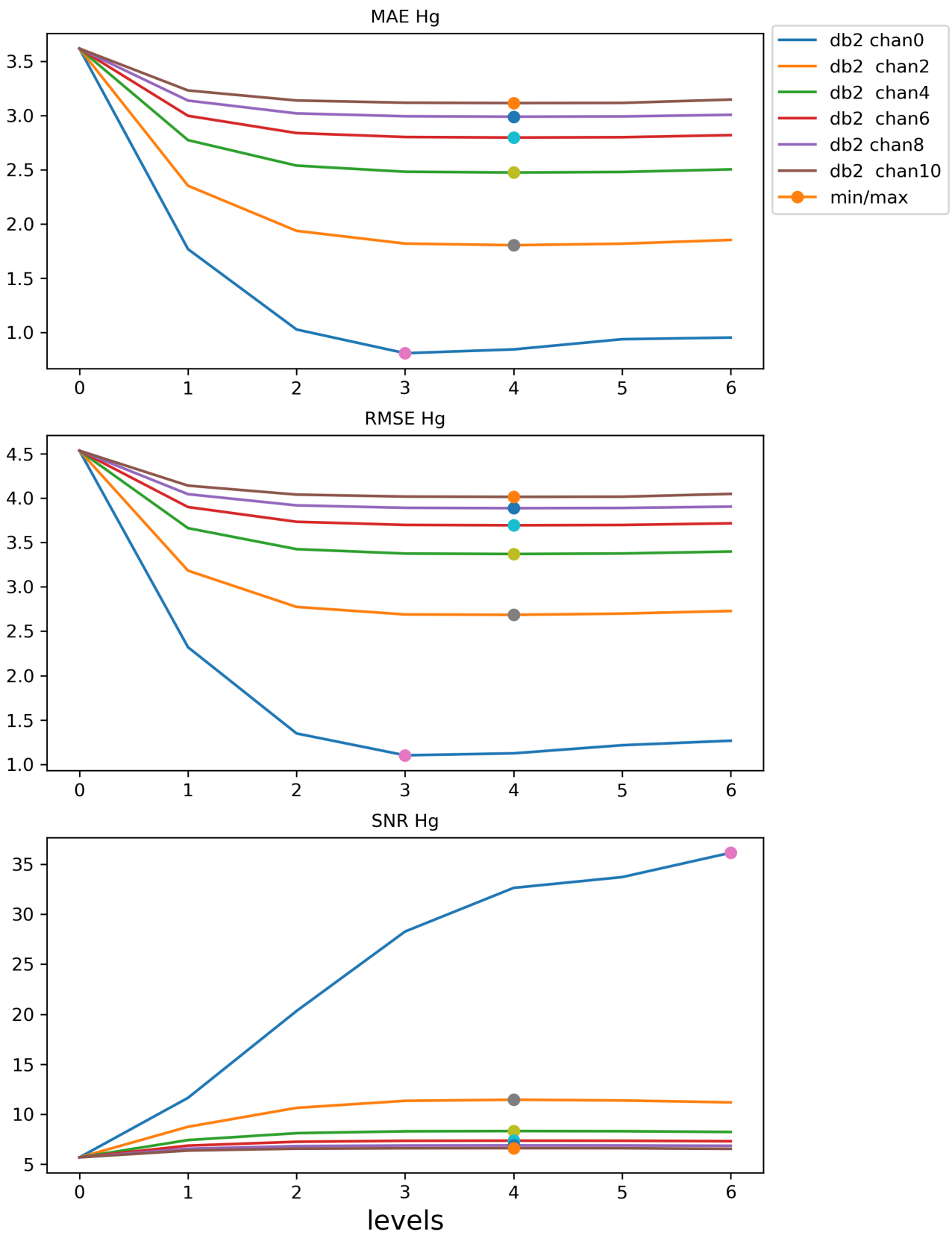


Figure 46: Average spatial performance quality of the Hg L3 Area for db2 at different s plotted against the number of decomposition levels

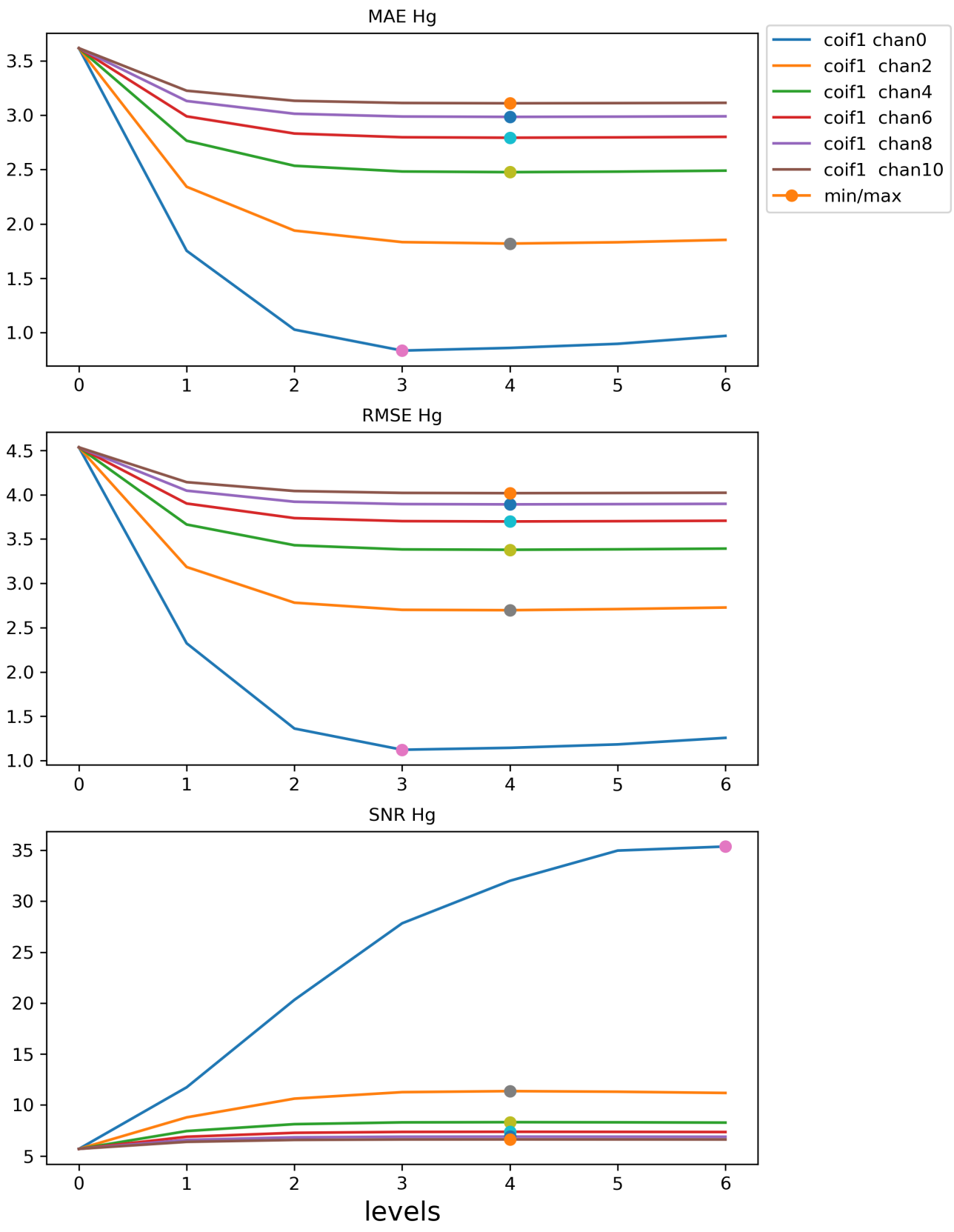


Figure 47: Average spatial performance quality of the Hg L3 Area for coif1 at different s plotted against the number of decomposition levels

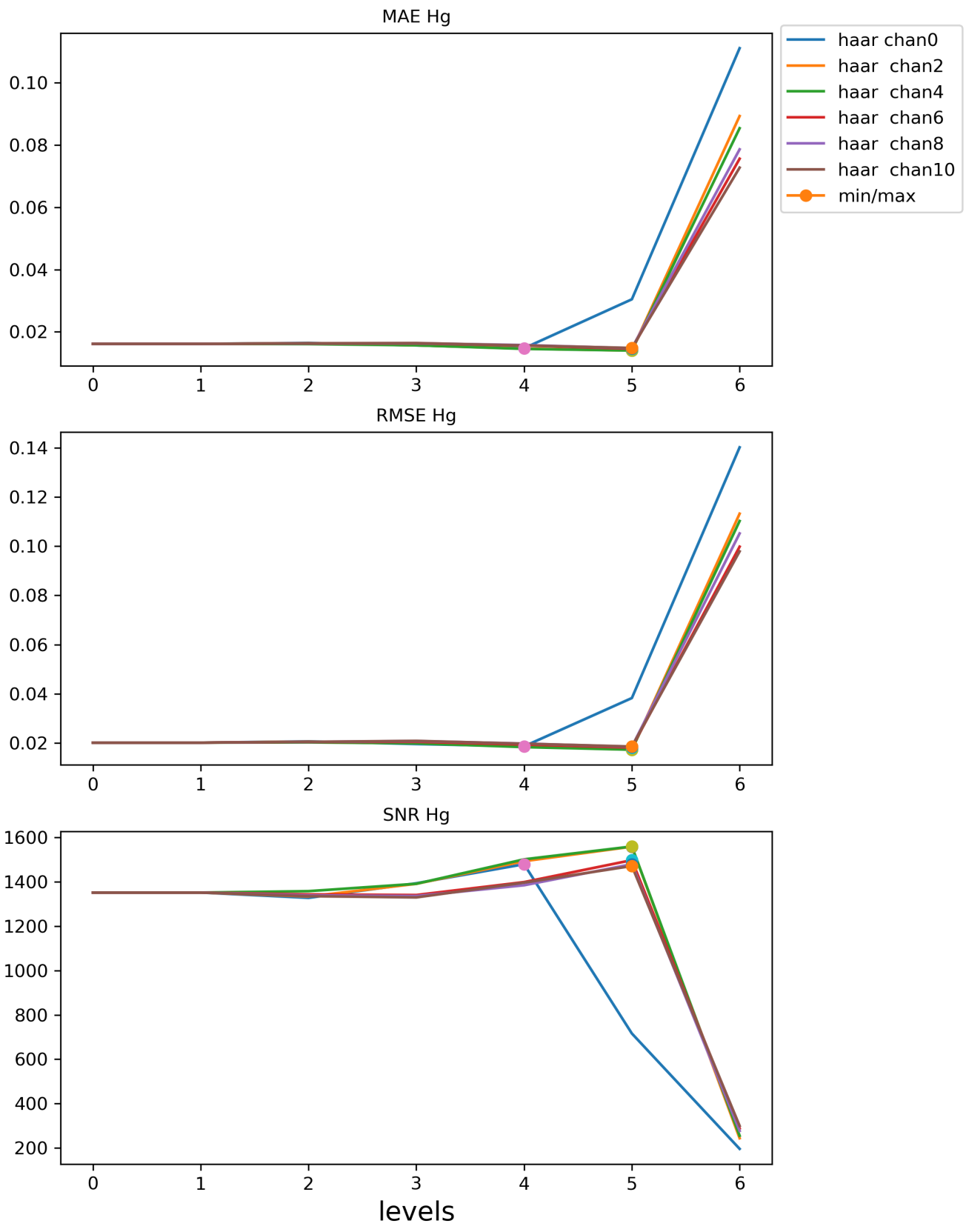


Figure 48: Summated spectral performance quality of the Hg L3 Area for haar at different s plotted against the number of decomposition levels

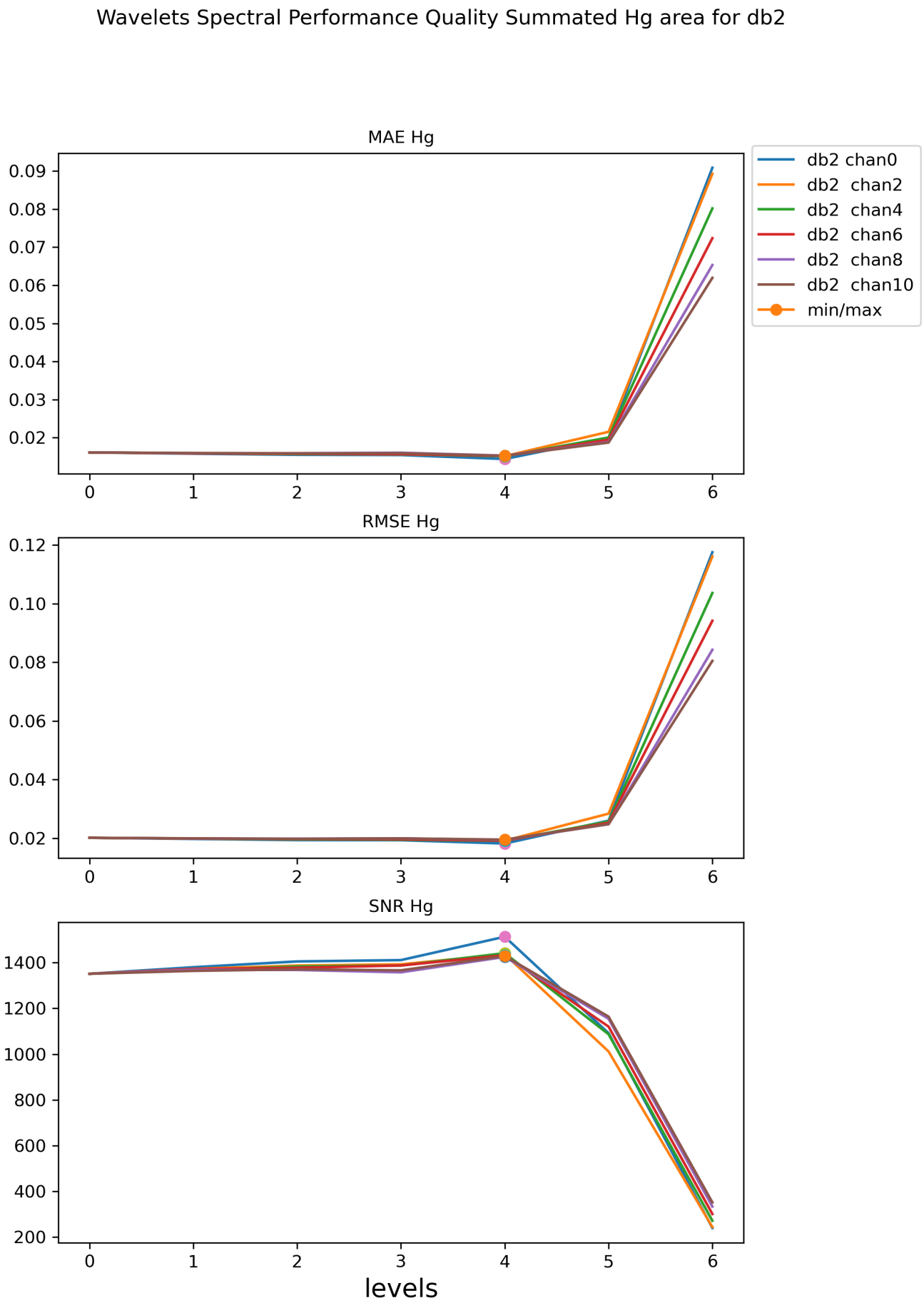


Figure 49: Summated spectral performance quality of the Hg L3 Area for db2 at different s plotted against the number of decomposition levels

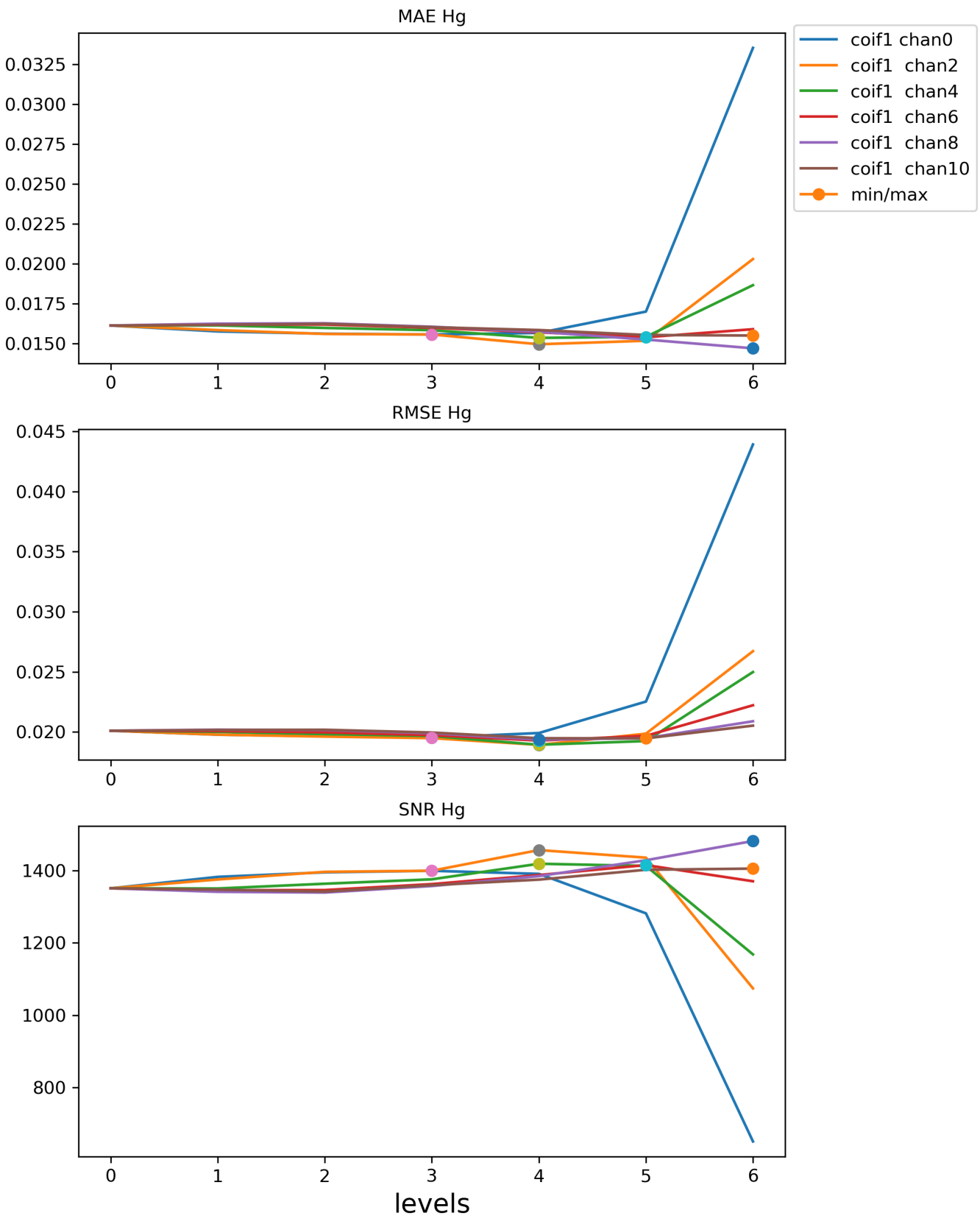


Figure 50: Summated spectral performance quality of the Hg L3 Area for coif1 at different s plotted against the number of decomposition levels

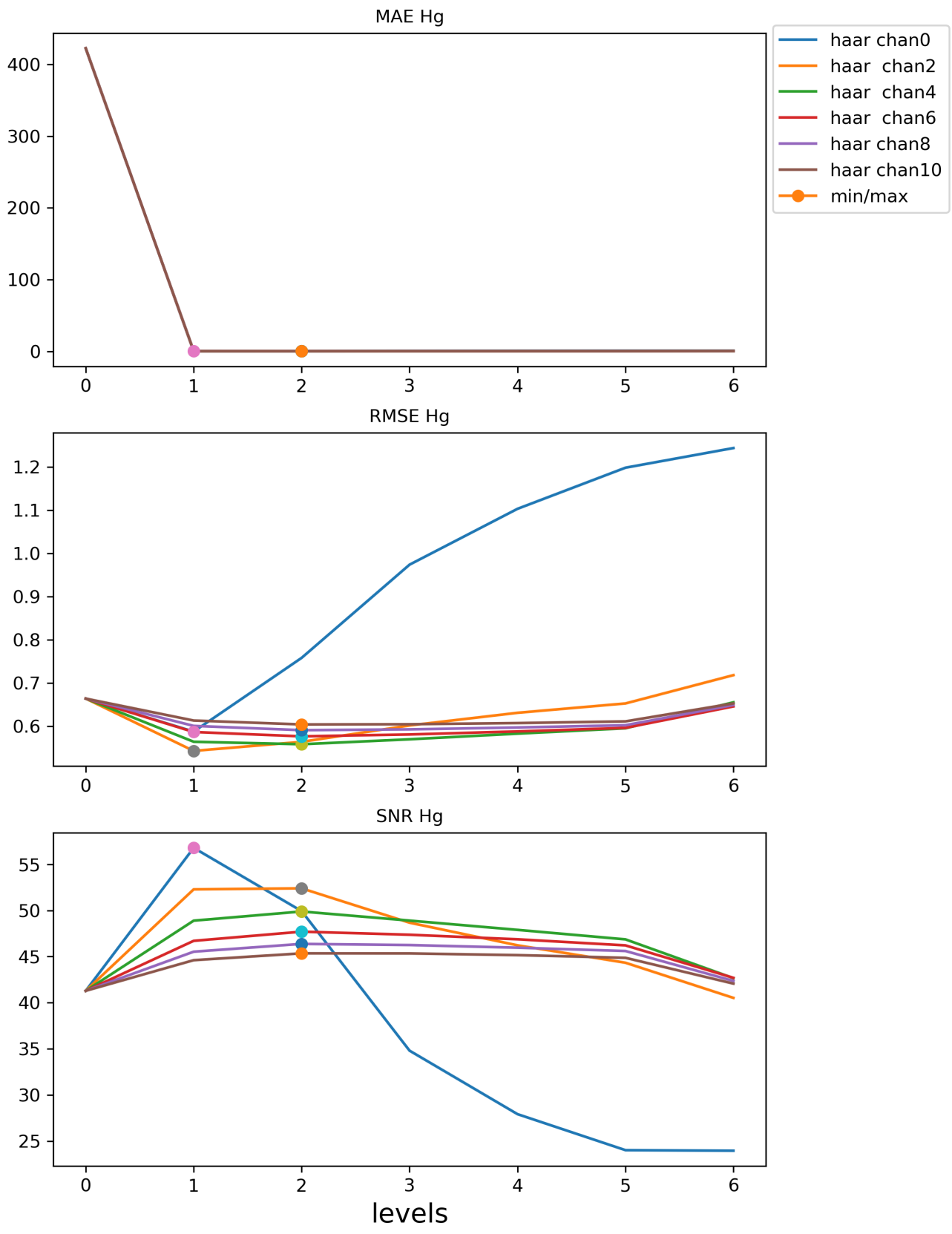


Figure 51: Summated spatial performance quality of the Hg L3 Area for haar at different s plotted against the number of decomposition levels

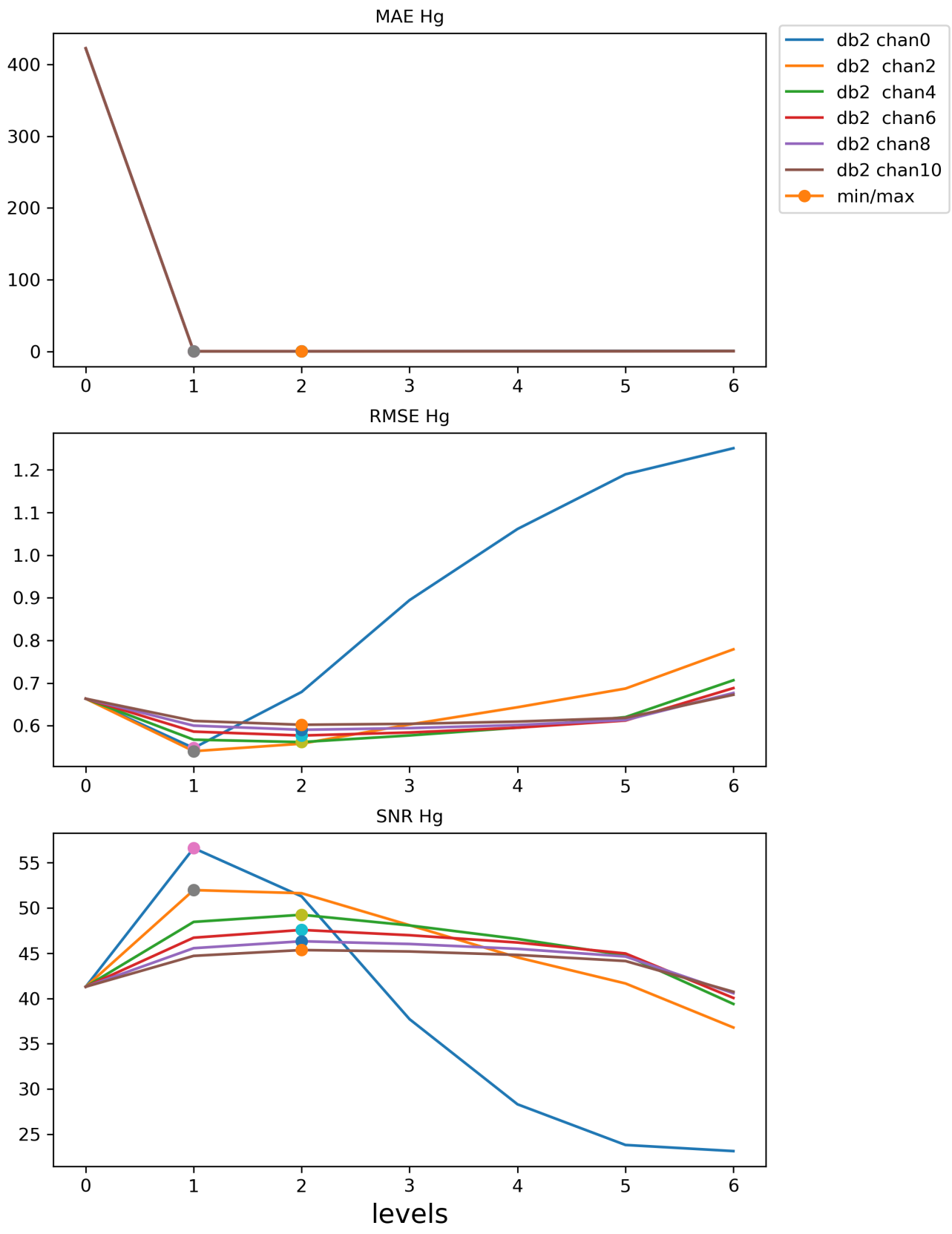


Figure 52: Summated spatial performance quality of the Hg L3 Area for db2 at different s plotted against the number of decomposition levels

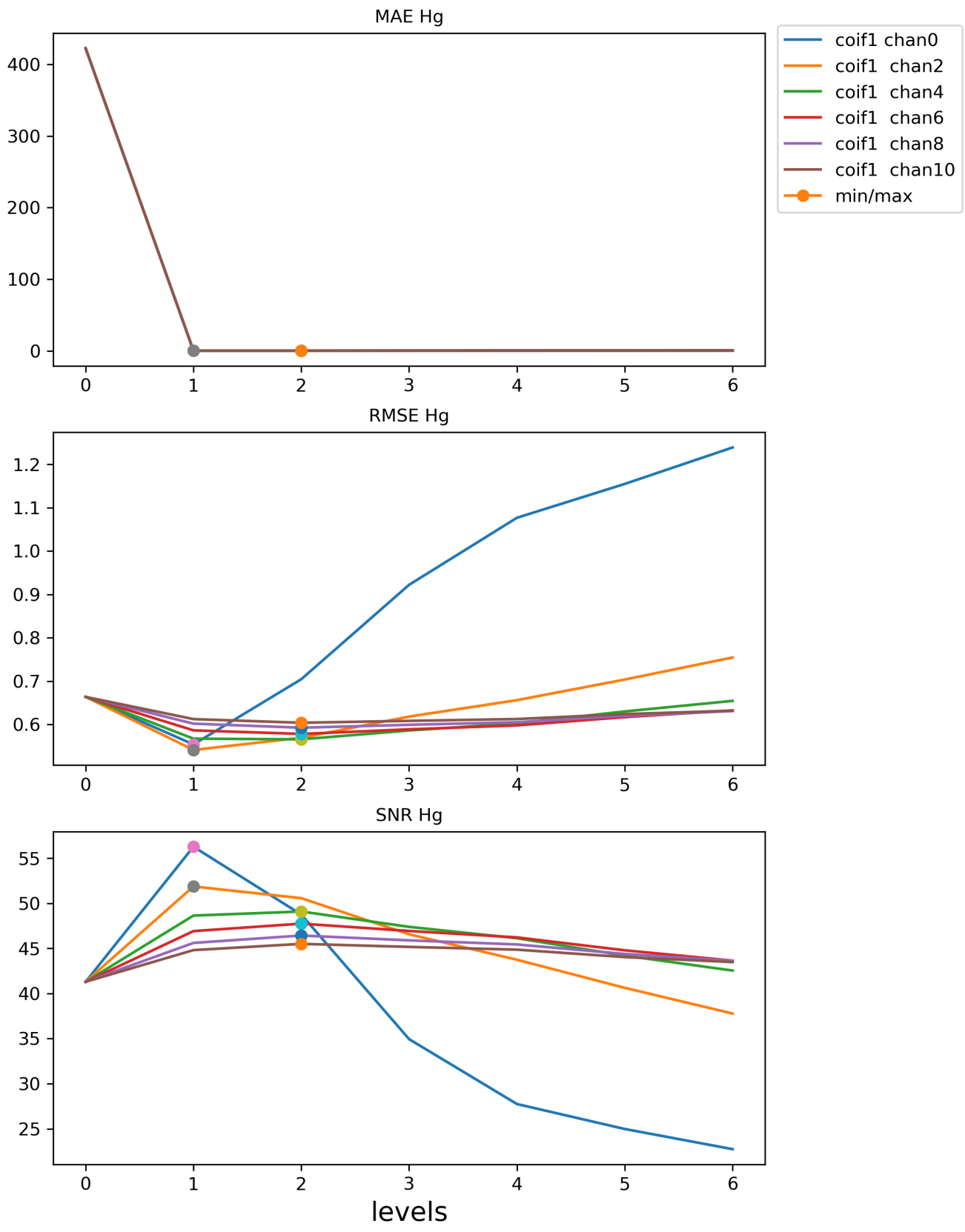


Figure 53: Summated spatial performance quality of the Hg L3 Area for coif1 at different s plotted against the number of decomposition levels

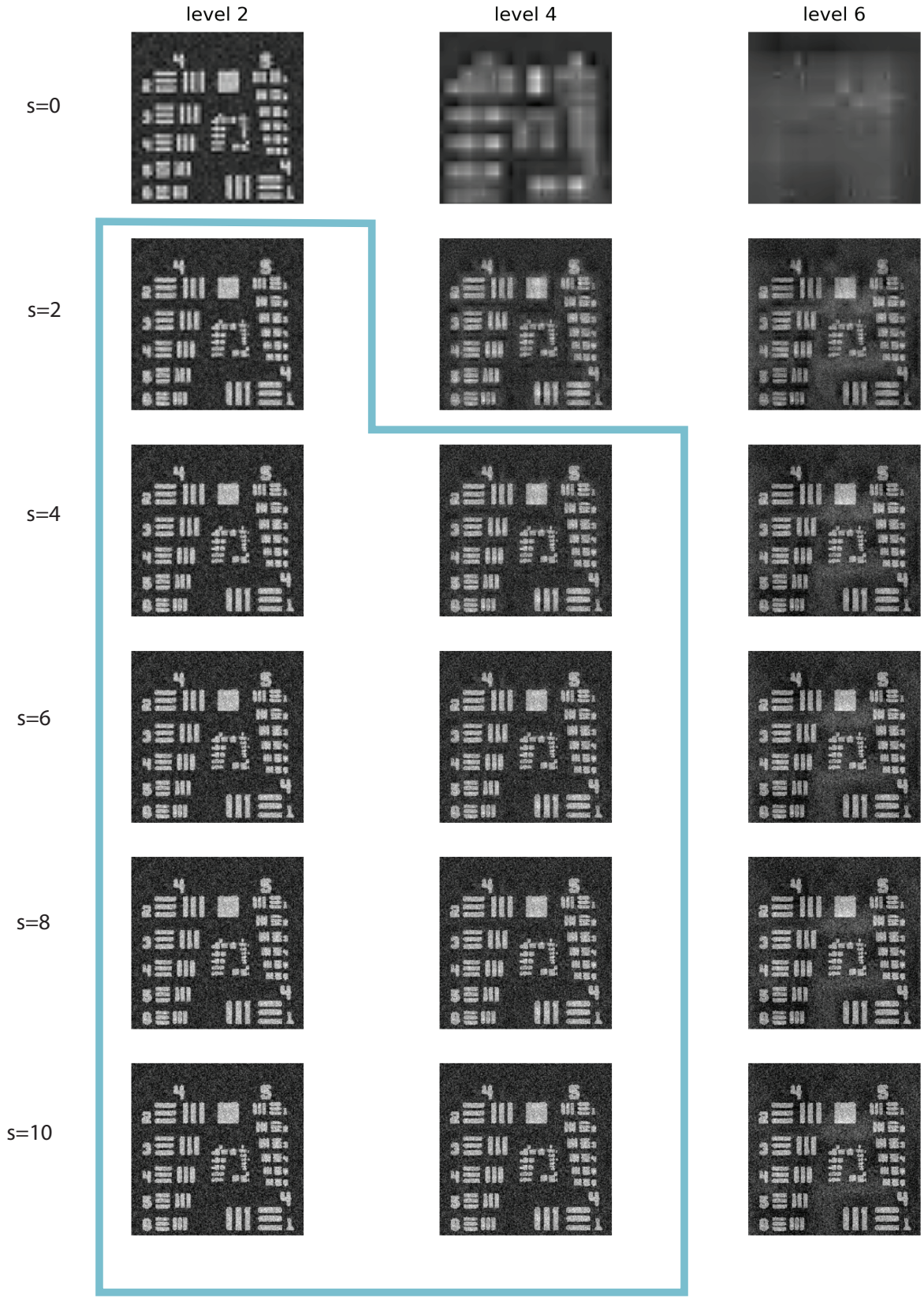


Figure 54: Peak Intensity maps of the db2 wavelet for combinations of number of neighbouring channels and levels

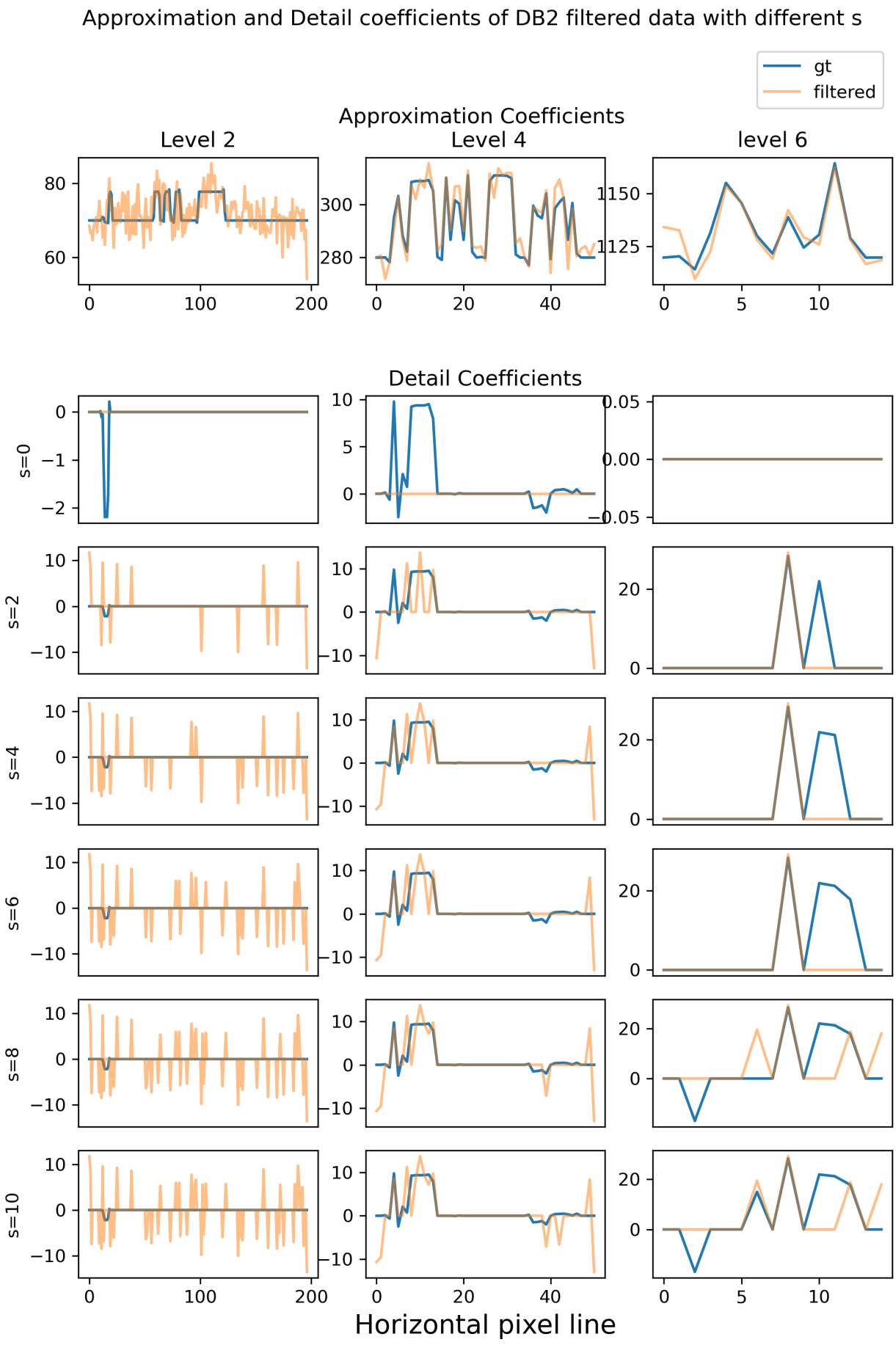


Figure 55: Approximation and Horizontal Detail Coefficients of a pixel row (see Figure 36) at channel 1080 obtained with db2 at level 2,4 and 6 for various s

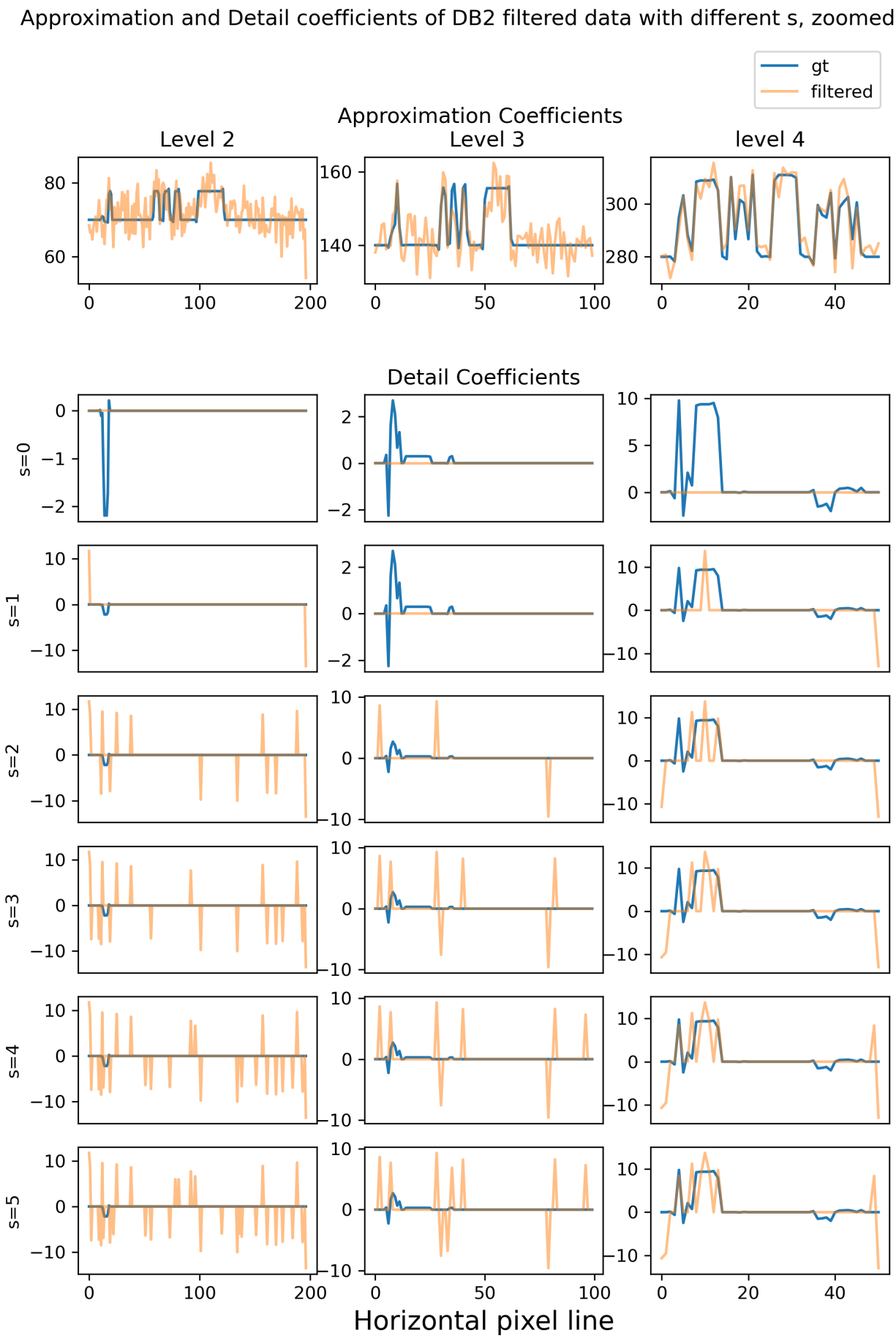


Figure 56: Approximation and Horizontal Detail Coefficients of a pixel row (see Figure 36) at channel 1080 obtained with db2 at level 2,3 and 4 for s between 0 and 5

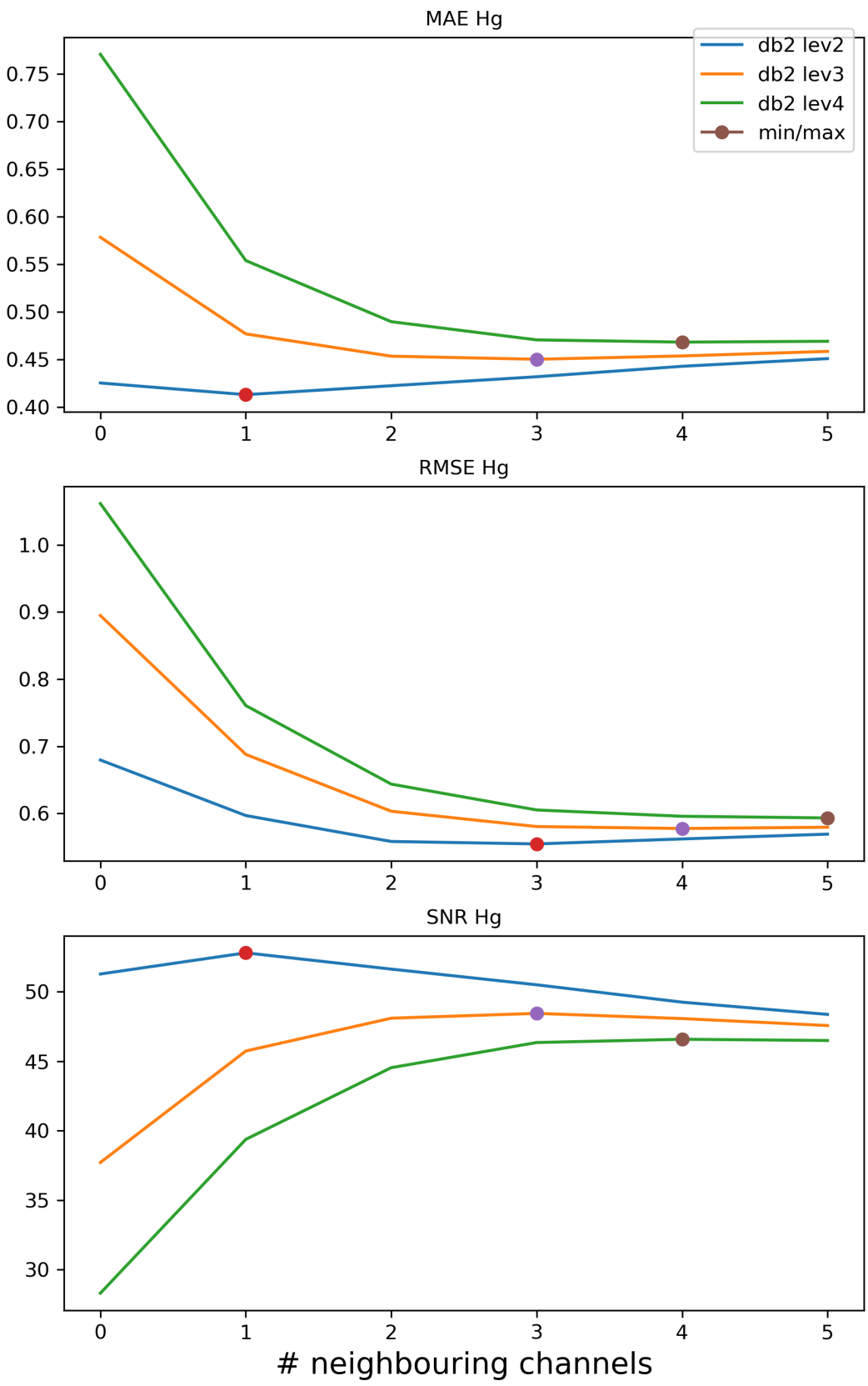


Figure 57: Summated spectral performance quality of the Hg L3 Area for DB2 at level 2,3, and 4 plotted against the number of neighbouring channels (s between 0 and 5)

Figure 58: Peak Intensity maps of db2 at decomposition level 2,3 and 4 for s between 0 and 5. The blue line surrounds the peak intensity maps that are largely free from spatial artefacting.

5.3.2 Discussion Testing Round 2b

In testing round 2, Figure 38 showed a clear level dependency for the optimal value for s. The higher the decomposition level, the higher the optimal value for s and vice versa. This can be explained by the fact that in low decomposition levels almost all detail coefficients represent noise, while in high decomposition levels it is more likely detail coefficients contain part of the signal and, as such, the threshold should be low enough not to discard that part of the signal. This is visible when looking at Figure 55, where we can compare the ground truth decomposition with the raw data decomposition coefficients. The higher the level, the more signal is found in the detail coefficients and the lower the s, the more detail is thresholded out.

When the average spectral and spatial performance quality was plotted for the Hg L3 area in Figure 43, Figure 42, Figure 44, Figure 45, Figure 46, and Figure 47, it could be seen that, despite the various number of neighbouring channels, almost every optimal performance was achieved at level 4. For the summated spectral and spatial performance quality plotted in Figure 48 Figure 49,Figure 50, Figure 51, Figure 52, and Figure 53, however, this was not the case. It could be seen that the summated spectral performance quality was better for somewhat higher levels than the summated spatial performance. A plausible reason is due to the fact that the data is denoised spatially and not spectrally. During higher-level decomposition, there is more redistribution of photons and more signal being present in the detail coefficients. Where this does change the spatial make-up of the signal, spectrally it comes down to mere averaging. Due to the perfect reconstruction principle of wavelet filter banks, namely, the amount of photons in every channel remains nearly constant. Another interesting observation was the difference between Figure 48, Figure 49 and Figure 50, where for the haar wavelet level 5 seemed optimal and for the db2 wavelet level 4 seemed optimal for all s, the coif1 wavelet had an s-dependend optimal level. This is most likely due to the support size of coif1 being larger. When the irregular MA-XRF data is filtered with a wavelet having larger support, more large detail coefficients are obtained. For lower-level decomposition, where mainly noise is present in the detail coefficients, it is then important to heighten the threshold, while for higher-level decomposition, where more signal is present in the detail coefficients, it becomes important to lower the threshold.

When looking at Figure 54, it becomes clear that there is not necessarily an optimal s or decomposition level, but rather an optimal combination of the two. It can, furthermore, be seen that both too low s and too high a level-decomposition are suboptimal. What does not necessarily become clear from Figure 54, but does become clear when regarding Figure 39, Figure 40, Figure 41, and Figure 55, is that the choice of s is like finding the balance between over-averaging (low s) and

barely removing noise (high s).

A clear optimal combination of s and the number of decomposition levels is hard to make. Despite level 4 performing best in the first testing round as well as the second testing round for the average performance quality, it could be seen that for the summated Hg L3 area this was not the case. Here, lower-level decomposition resulted in higher spatial performance quality, while higher-level decomposition resulted in higher spectral performance quality. It is important to recall that this higher-level decomposition caused spatial artefacting. Furthermore, the significant performance quality improvements were generally until level 3. Similarly, for the neighbouring channels, major improvements were made by increasing the number of neighbouring channels from $s = 0$ to $s = 5$, but after that, between $s = 5$ and $s = 10$, no significant improvements were made and more noise was contained in the data.

Therefore, an optimal level/neighbouring channel combination was narrowed down to be between decomposition levels 2 and 4 and between $s = 0$ and $s = 5$. In Figure 56 it could be seen that for level 3, the approximation coefficients of the ground truth were in accordance with the spatial pattern of the synthetic dataset and noise was effectively removed when comparing the coefficients to the approximation coefficients of level 2. when looking at Figure 57 a suitable number of neighbouring channels seems to be at $s = 3$. Also, the peak intensity maps in Figure 58 show visually well-performing results at the level 3 $s = 3$ combination.

Due to the support size of the coif1 wavelet resulting in more high detail coefficients, we will continue the third testing round with the haar and db2 wavelet. Furthermore, a suitable combination of a level-3 decomposition with $s = 3$ will be used.

5.4 Testing Round 3

From previous testing rounds, we obtained a suitable combination of filter bank denoising parameters. For the mother wavelet, this resulted in two good design choices: the haar and db2 wavelet. For the amount of decomposition levels, 3 was found satisfactory. Furthermore, we determined a proper value for s to equal 3. We will proceed to refer to the two different filters as 'haar lev3 s3' and 'db2 lev3 s3'. These filter designs are now compared to two median filters, one with a 3×3 kernel and the other with a 5×5 kernel. Moreover, the synthetic dataset has been simulated with three different degrees of noise, in other words, three different well times – of 1, 0.75, and 0.5 seconds – on which the filters are tested.

5.4.1 Observations Testing Round 3

Explaining and Observing the Performance Quality Sheet. Following, the performance quality of the raw data, the median filters, the haar lev3 s3, and the db2 lev3 s3 are plotted. Every sheet has a similar make-up. In the top row, from left to right the ground truth Hg L3 peak intensity map is shown (as a reference), next to that is the peak intensity map of the filtered data. At the top right corner, there is a small table where the performance quality of the summated data is presented. Here, the spectral sum performance quality is determined by first summing all the spectra of the filtered data and then comparing that to the sum spectrum of the ground truth data. Similarly, the spatial sum performance quality is determined by first summing all the channels of the filtered data and then, comparing that to the sum of channels of the ground truth data.

In the leftmost column, the spectral performance quality is plotted for every pixel. Similarly, the spatial performance is plotted for every channel in the middle column, giving an overview of the spectral performance quality per pixel and the spatial performance quality per channel, respectively. In the right column, the spectral performance for the mercury L3 peak is plotted, based on 50 channels covering the peak. Above all nine performance quality plots, the average performance is indicated (AVG). For the spectral performance quality plots also the standard deviation is indicated (SIG).

The raw data is plotted in Figure 62, the 3×3 median filtered data is plotted in Figure 63, the 5×5 median filtered data is plotted in Figure 64, the haar lev4 s5 filtered data is plotted in Figure 65, and the db2 lev4 s5 filter data is plotted in Figure 66. When regarding the figures, it can be seen that the values for the wavelet-filtered data lie close to one another. For the average spatial and spectral performance quality, both median and wavelet-filtered data obtain better results than when the data is not filtered. The median filtered data, however, has better average performance quality than the wavelet filtered data. Yet, the summated

spatial and spectral performance quality is significantly worse for the median filtered data than for the raw and wavelet filtered data. Moreover, the median 5x5 filtered data has a RMSE of the spectral sum significantly lower than that of the wavelet-filtered and median 3x3 filtered data.

The standard deviation for the MAE and RMSE of the spectra of the median filtered data is smaller than that of the wavelet filtered data. Still, both wavelet-filtered data and median-filtered data have a smaller standard deviation than the raw shot noise-polluted data. Both the average SNR and the standard deviation of SNR become bigger when filtered.

A visual inspection of the spectral performance quality of the Hg L3 area of the median filters shows the heightening of the errors at the edges of intensity changes. This is more prominent in the 5×5 median filtered data than the 3×3 median filtered data. In Figure 64 the Hg L3 peak intensity map even becomes visible when plotting the overview for the entire dataset (lower left) of the spectral SNR.

When comparing the raw, shot noise-polluted data from Figure 62 with the haar lev3 s3 filtered data from Figure 65, the wavelet filter generally results in data being closer to the ground truth. However, the summated spectral performance is higher for the raw data.

Also, the performance quality of filters for a higher amount of noise is determined. The performance is visualised for the median 3×3 filter in Figure 68 and Figure 72, the haar lev3 s3 filter in Figure 69 and Figure 73, and the db2 lev3 s3 filter in Figure 70 and Figure 74. For reference, the raw shot noise-polluted data is shown in Figure 67 and Figure 71. It can be seen that the performance quality is worse than in the case of a higher dwell time. The average performance quality of the median filter is still better than the average performance quality of the wavelets. On the other hand, the sum performance quality of the wavelet filters is also still better than that of the median filter.

Plotting the Sum Spectra. When plotting the sum spectra of the ground truth data, the raw data, the median filtered data and the wavelet filtered data in Figure 59 it can be seen that the wavelet filtered data and raw data follow the ground truth sum spectrum and the 3x3 median filtered data diverges from the ground truth. When increasing the amount of noise, the sum spectra visible inFigure 60 and Figure 61 are obtained. The wavelets still follow the ground truth sum spectrum, while the median filtered data performs increasingly worse than in Figure 59.

Plotting the Performance Quality against Various amounts of Noise. In Figure 75, it can be seen that the higher the amount of noise, the lower the

performance quality, both for the median filter and the wavelet filters. It can be seen that the median filter has the best average performance quality, while it has the worst sum performance quality. Moreover, the spectral sum performance quality of the wavelet filter is nearly identical to that of the raw data. Both the wavelet and the median filter succeed in reducing the level of noise on average. But only the wavelet has a good sum performance quality as well.

The purple dotted line is plotted to show what dwell time is necessary for the raw data to achieve the same performance quality as the filtered data. What can be seen is that the performance quality of the wavelet-filtered data obtained with a dwell time of 0.5 seconds is approximately equal to unfiltered data obtained with a dwell time of 0.75 seconds.

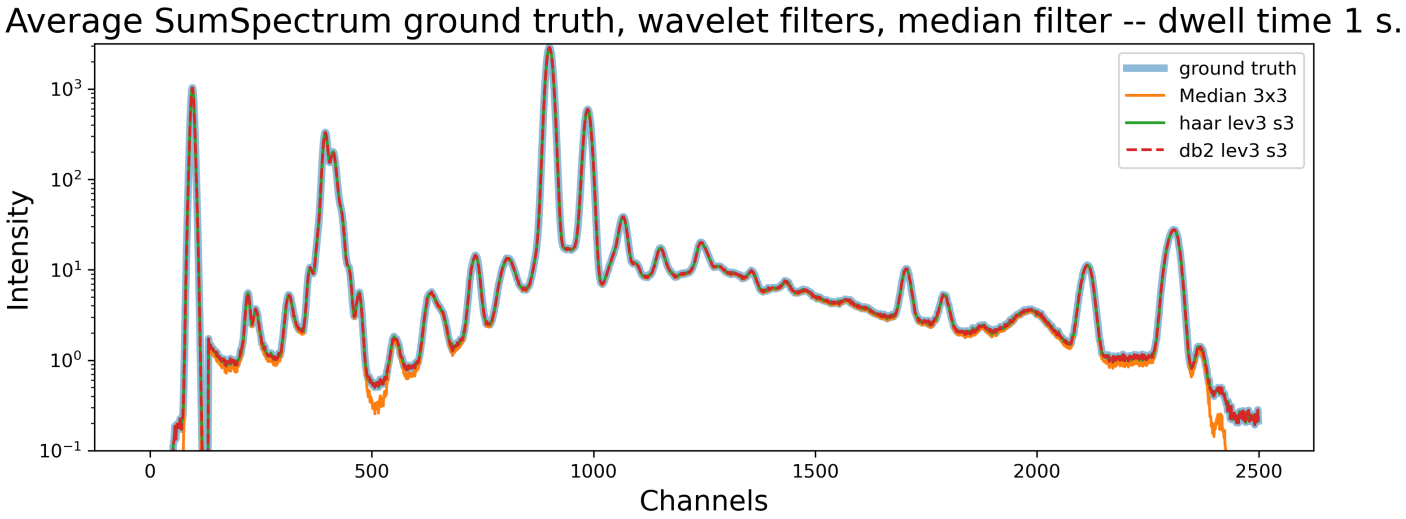


Figure 59: Averaged Sum Spectrum of Ground Truth, 3x3 Median filter, haar lev3 s3 filter and db2 lev3 s3 filter for a dwell time of 1 s.

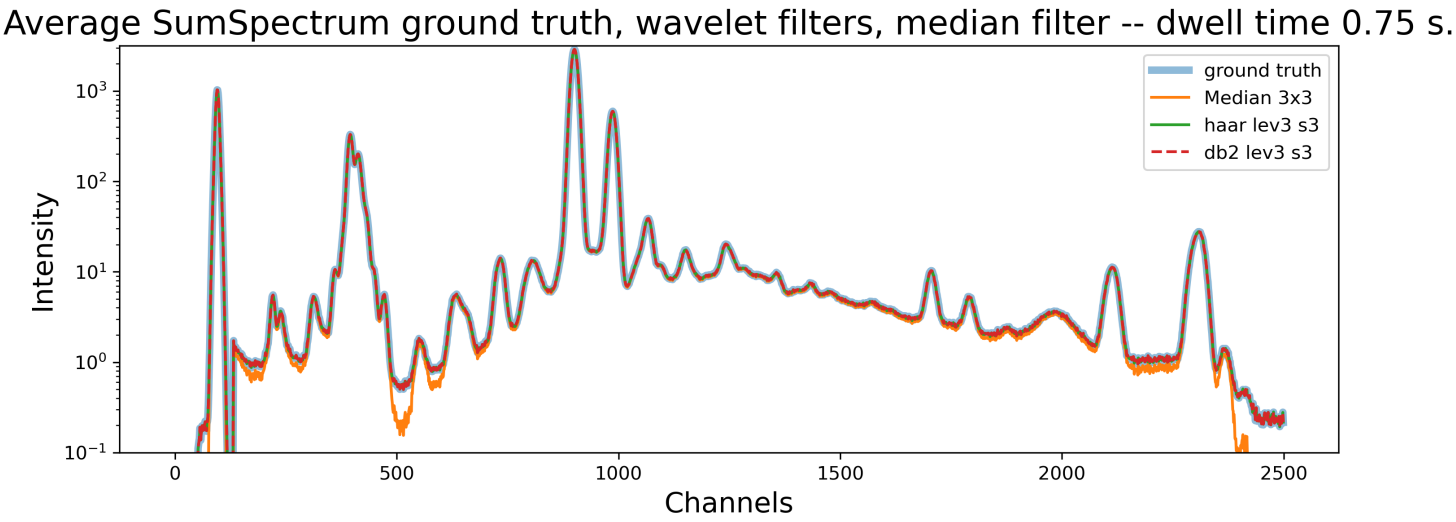


Figure 60: Averaged Sum Spectrum of Ground Truth,3x3 Median filter, haar lev3 s3 filter and db2 lev3 s3 filter for a dwell time of 0.75 s.

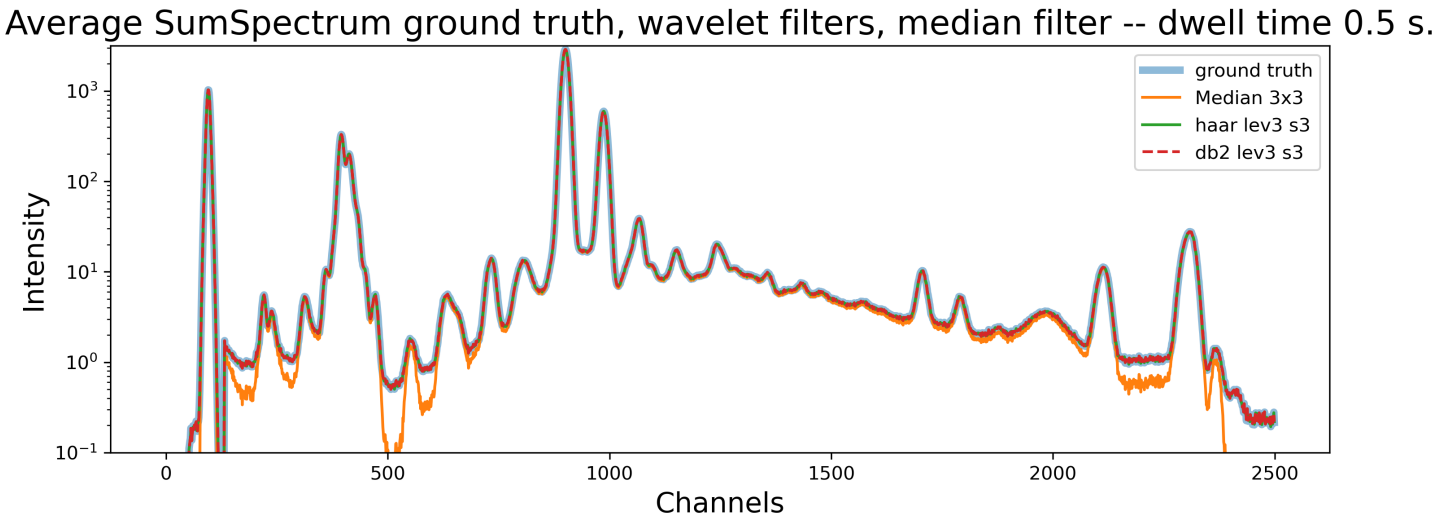
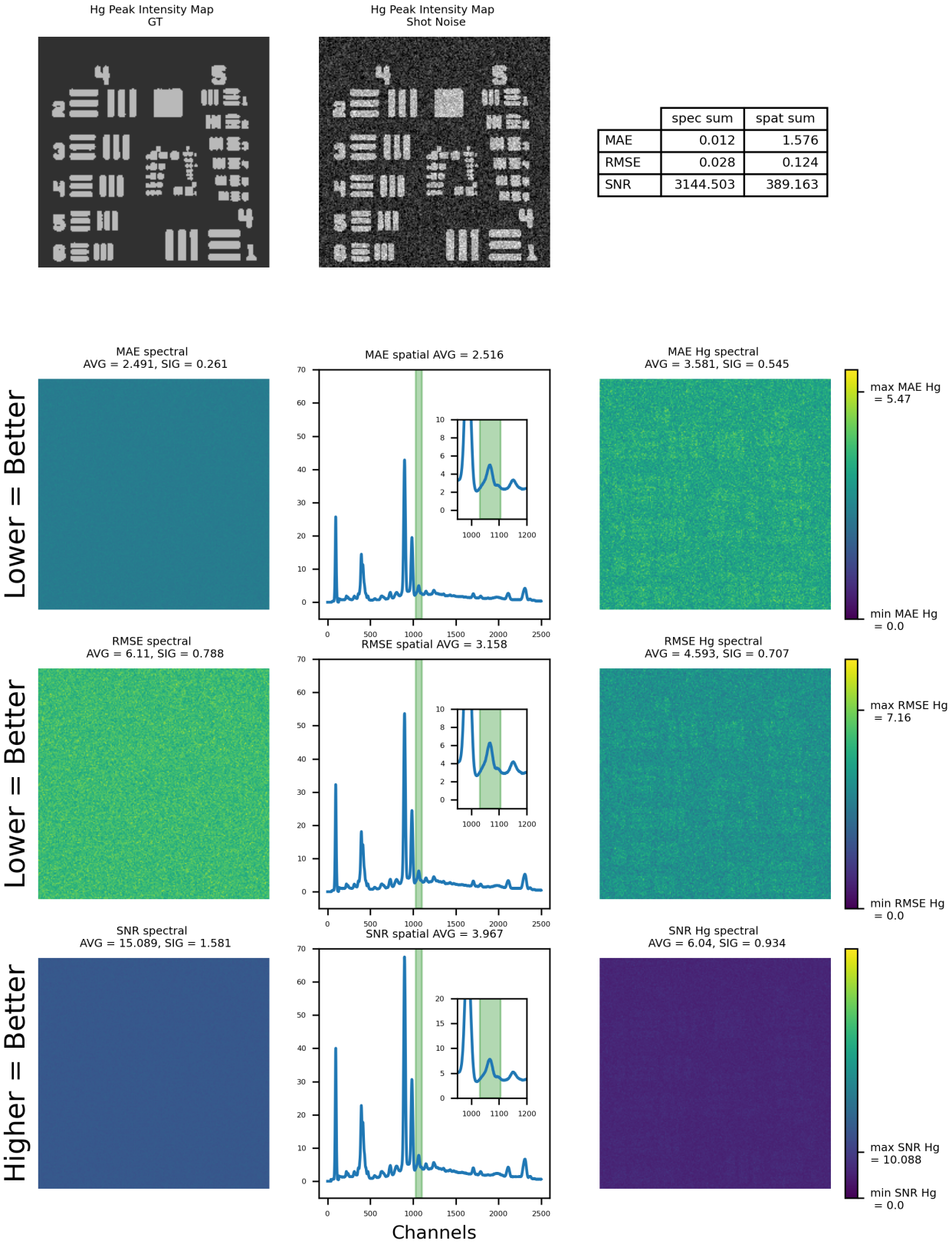


Figure 61: Averaged Sum Spectrum of Ground Truth, 3x3 Median filter, haar lev3 s3 filter and db2 lev3 s3 filter for a dwell time of 0.5 s.

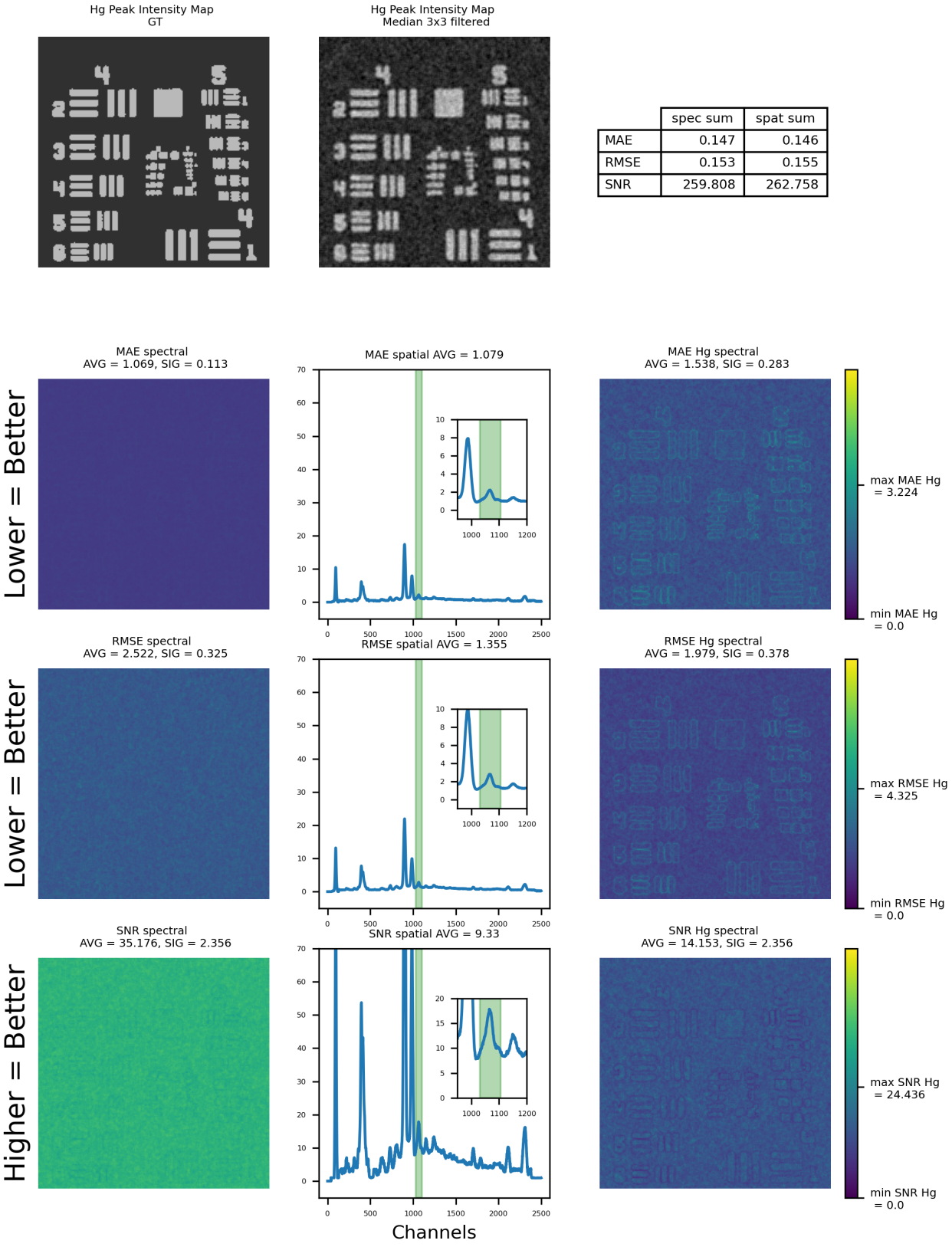
Shot Noise-Poluted Data with a dwell time of 1 sec.



	spec sum	spat sum
MAE	0.012	1.576
RMSE	0.028	0.124
SNR	3144.503	389.163

Figure 62: Performance Quality of Shot Noise-Polluted Data with a Dwell Time of 1 s.

Median 3x3 filtered data with a dwell time of 1 sec.



	spec sum	spat sum
MAE	0.147	0.146
RMSE	0.153	0.155
SNR	259.808	262.758

Figure 63: Performance Quality of 3x3 Median Filtered Data with a Dwell Time of 1 s.

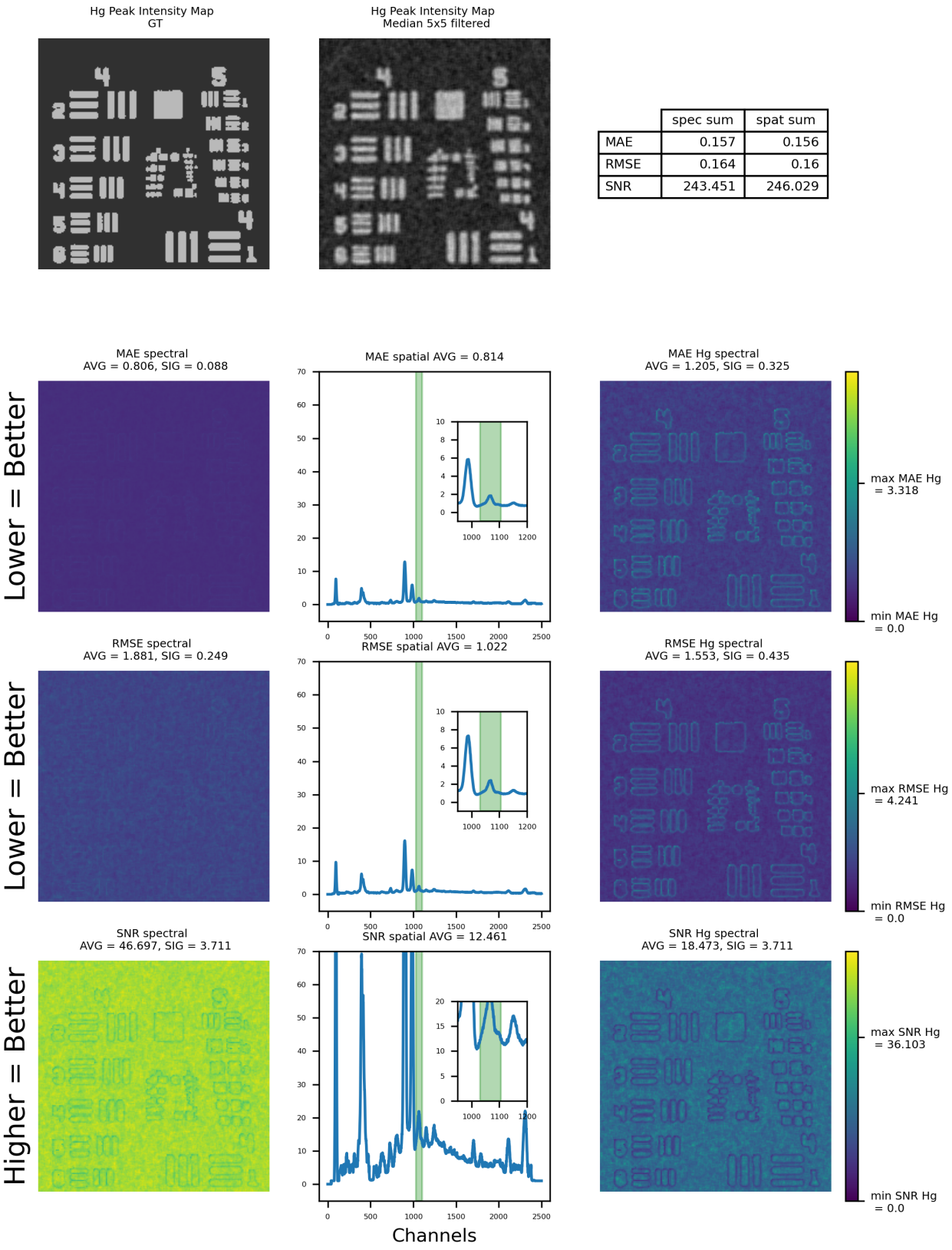


Figure 64: Performance Quality of 5x5 Median Filtered Data with a Dwell Time of 1 s.

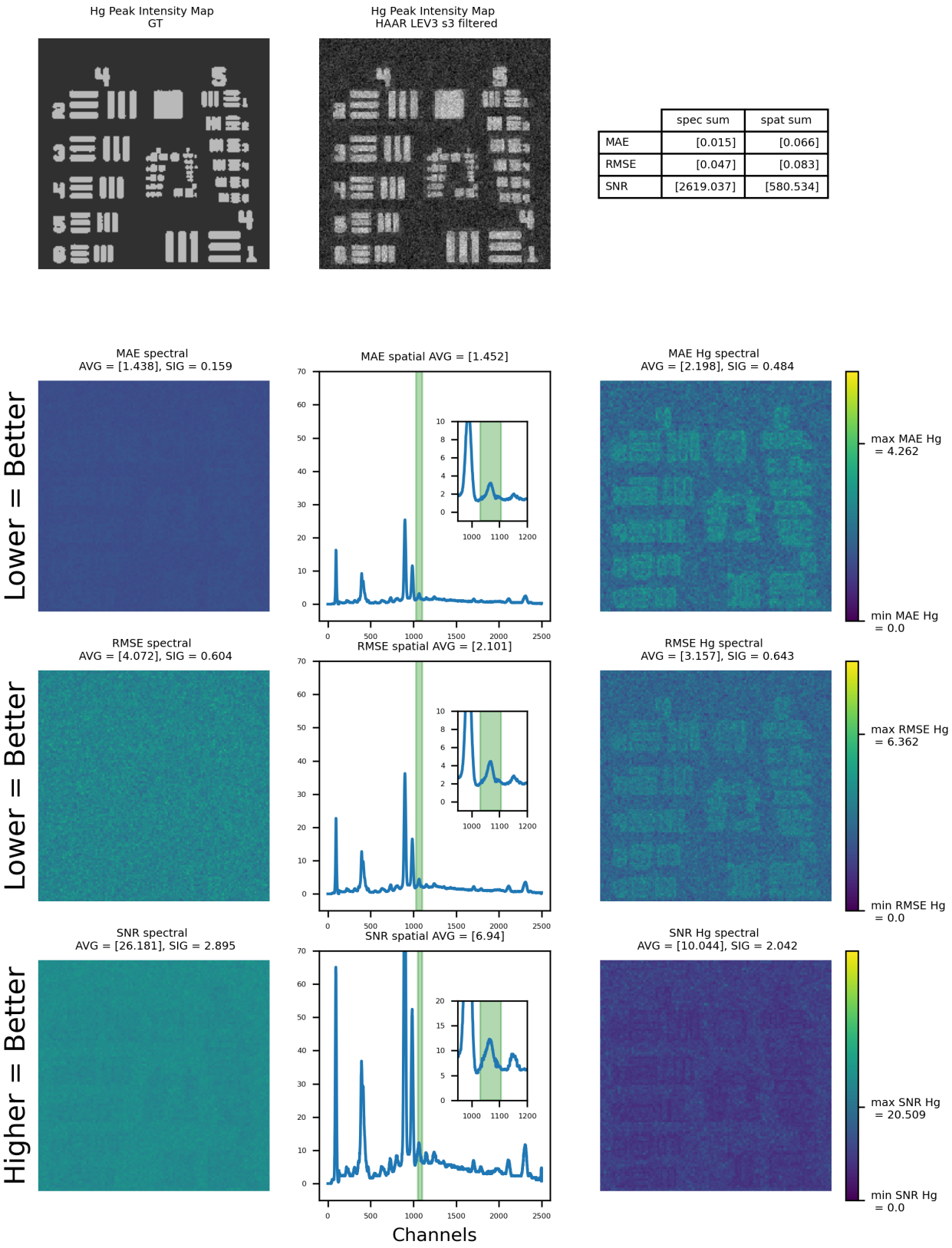


Figure 65: Performance Quality of haar Lev3 s3 Filtered Data with a dwell time of 1 s.

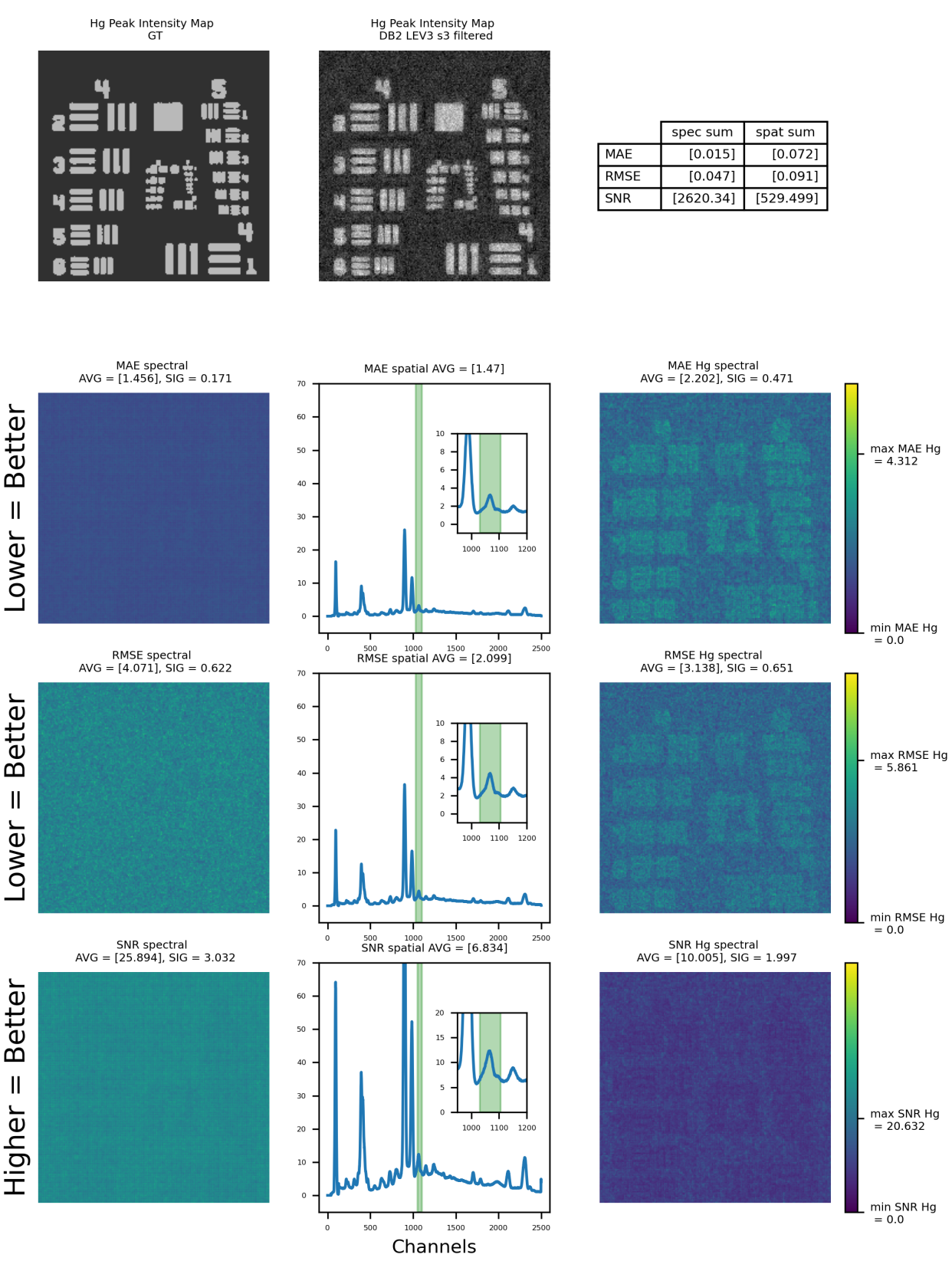


Figure 66: Performance Quality of db2 Lev3 s3 Filtered Data with a dwell time of 1 s.

Shot Noise-Poluted Data with a dwell time of 0.75 sec.

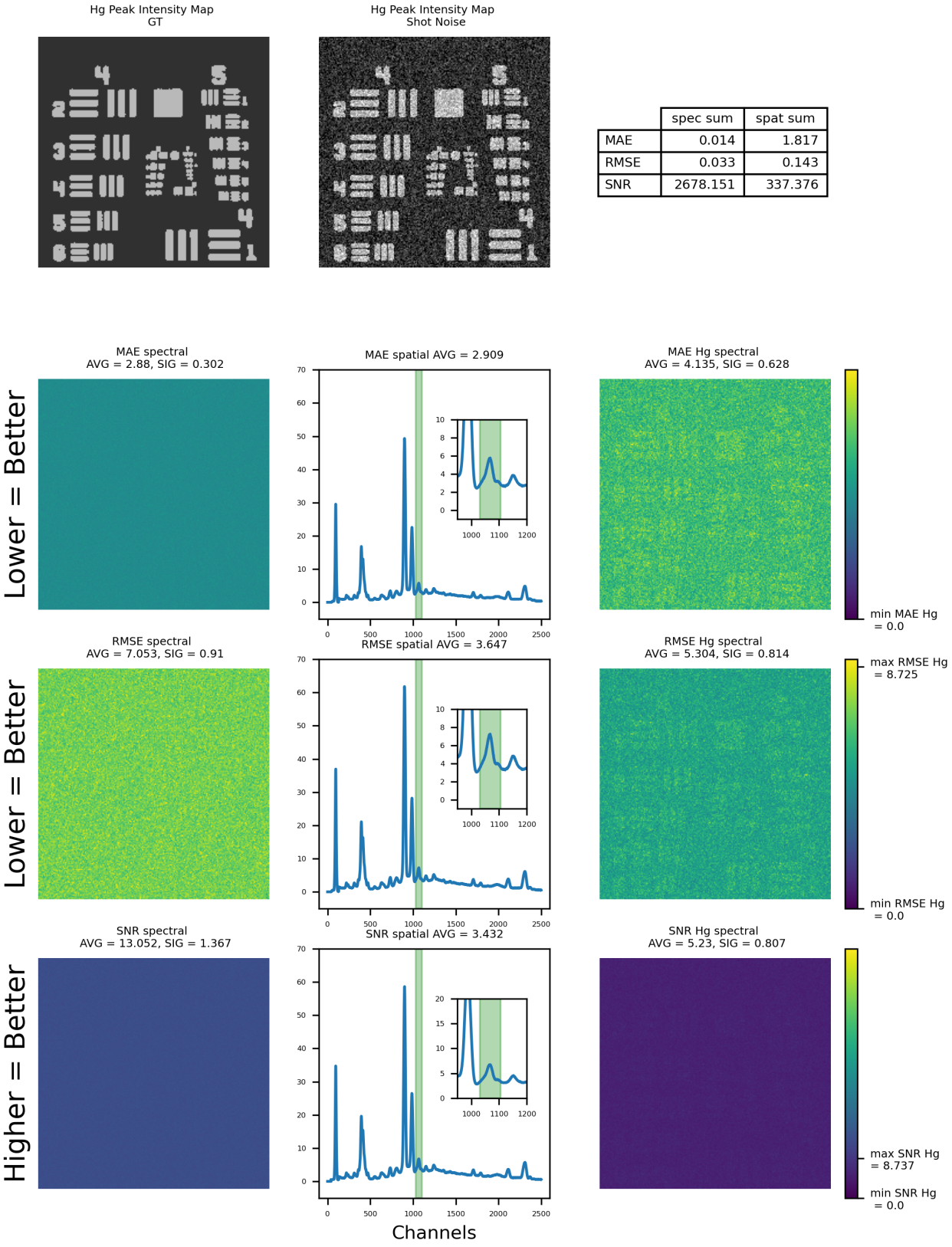


Figure 67: Performance Quality of Shot Noise-Polluted Data with a Dwell Time of 0.75 s.

Median 3x3 filtered data with a dwell time of 0.75 sec.

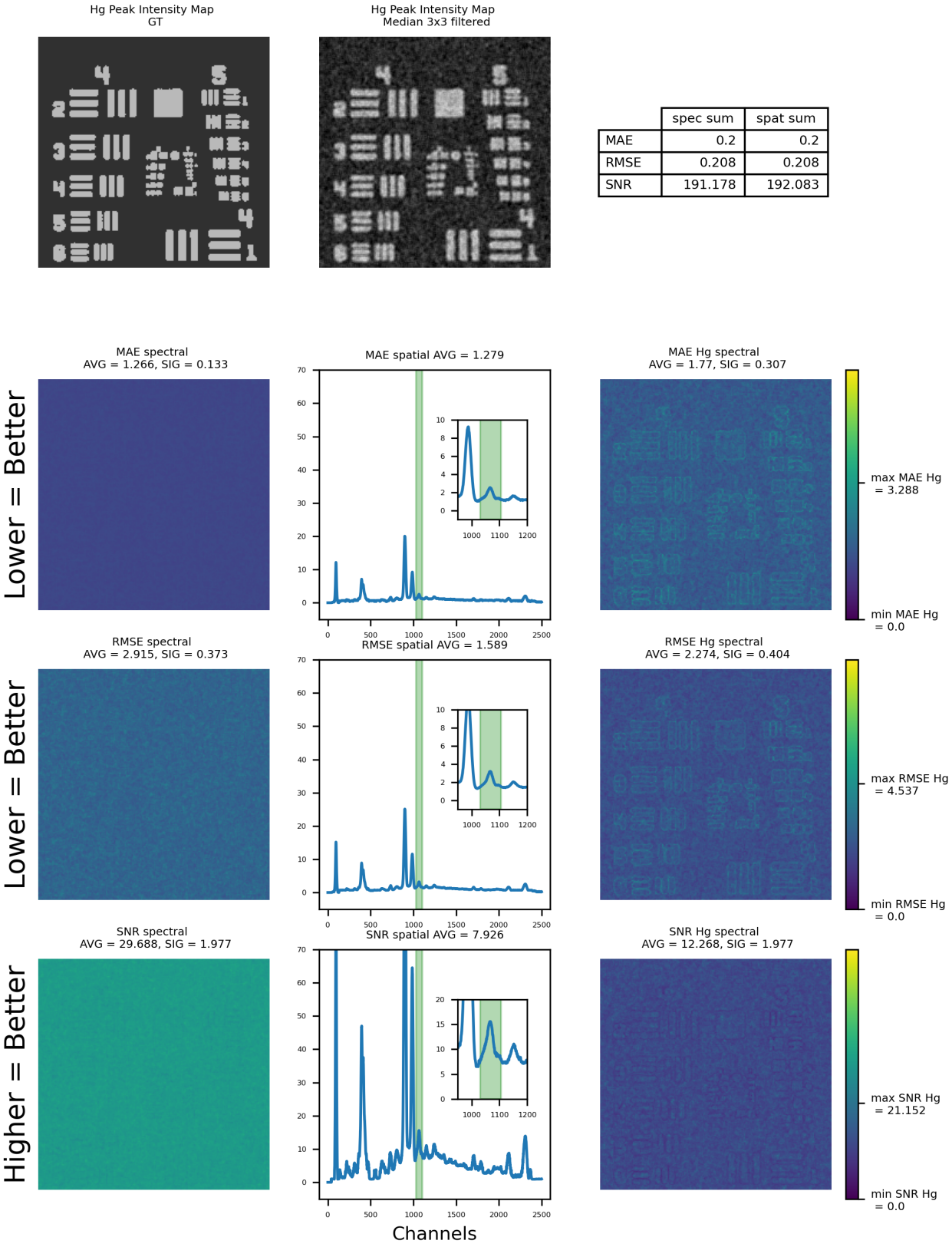


Figure 68: Performance Quality of 3x3 Median Filtered Data with a Dwell Time of 0.75 s.

HAAR LEV3 s3 filtered data with a dwell time of 0.75 sec.

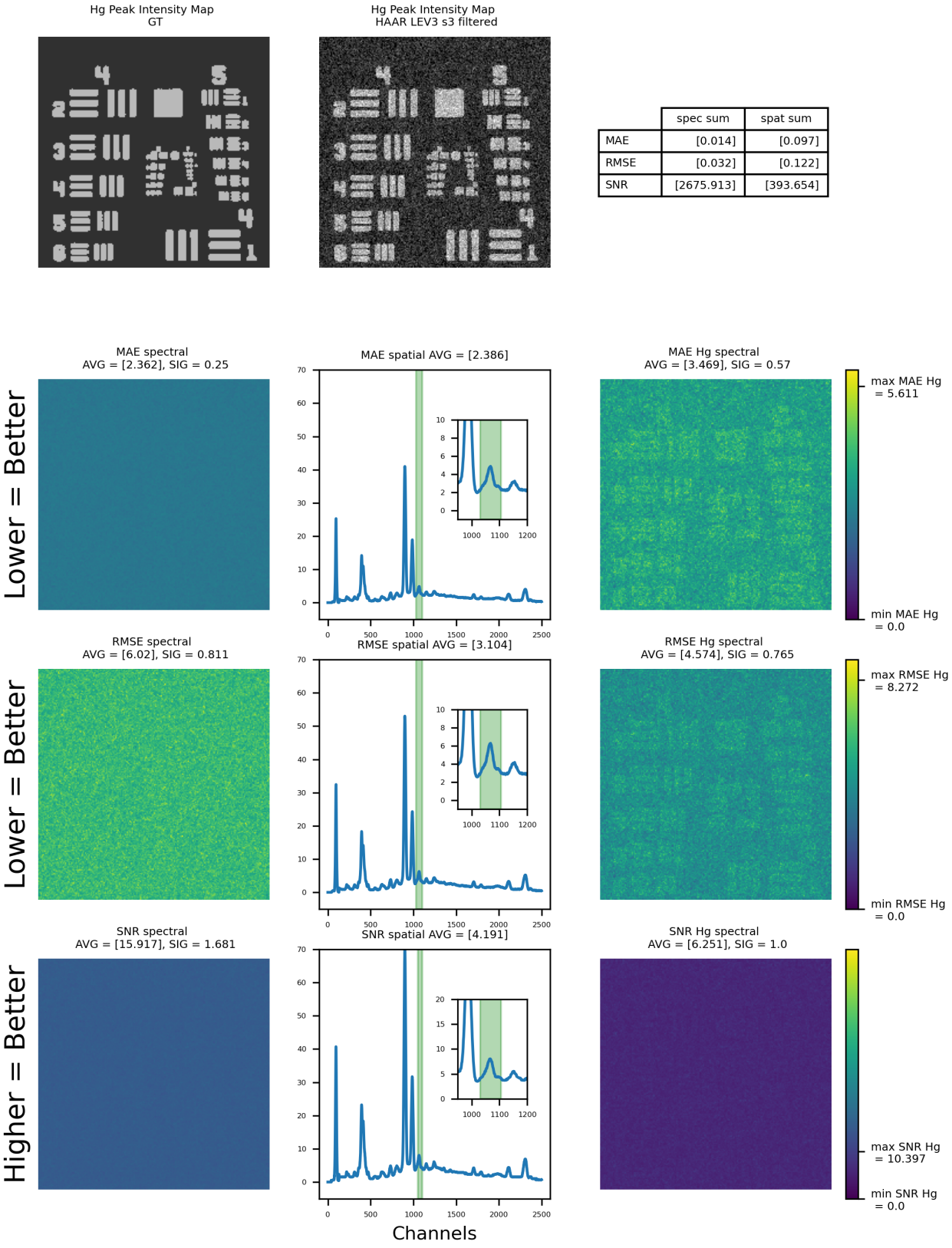


Figure 69: Performance Quality of Haar Lev3 s3 Filtered Data with a dwell time of 0.75 s.

DB2 LEV3 s3 filtered data with a dwell time of 0.75 sec.

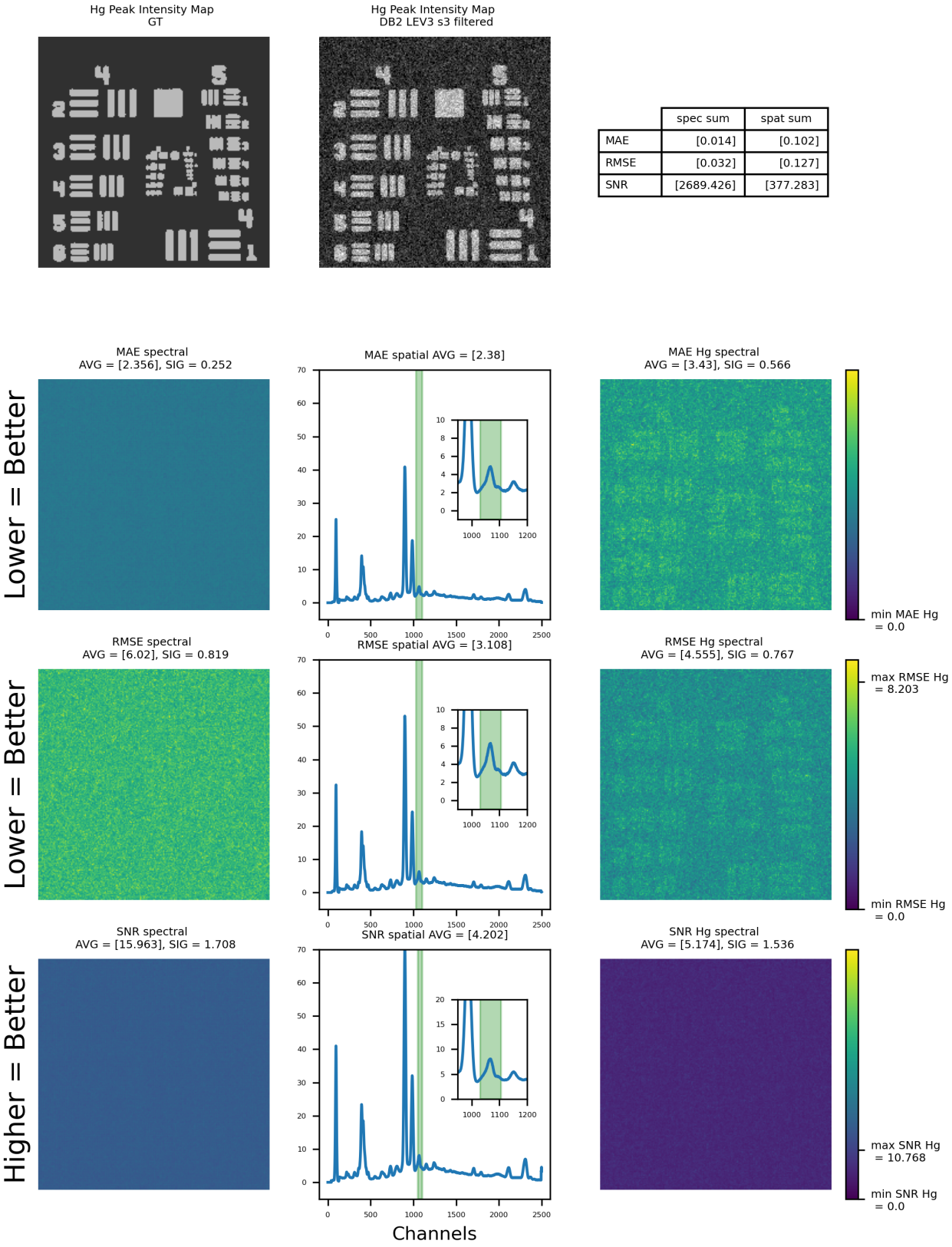


Figure 70: Performance Quality of db2 Lev 3 s3 Filtered Data with a Dwell Time of 0.75 s.

Shot Noise-Poluted Data with a dwell time of 0.5 sec.

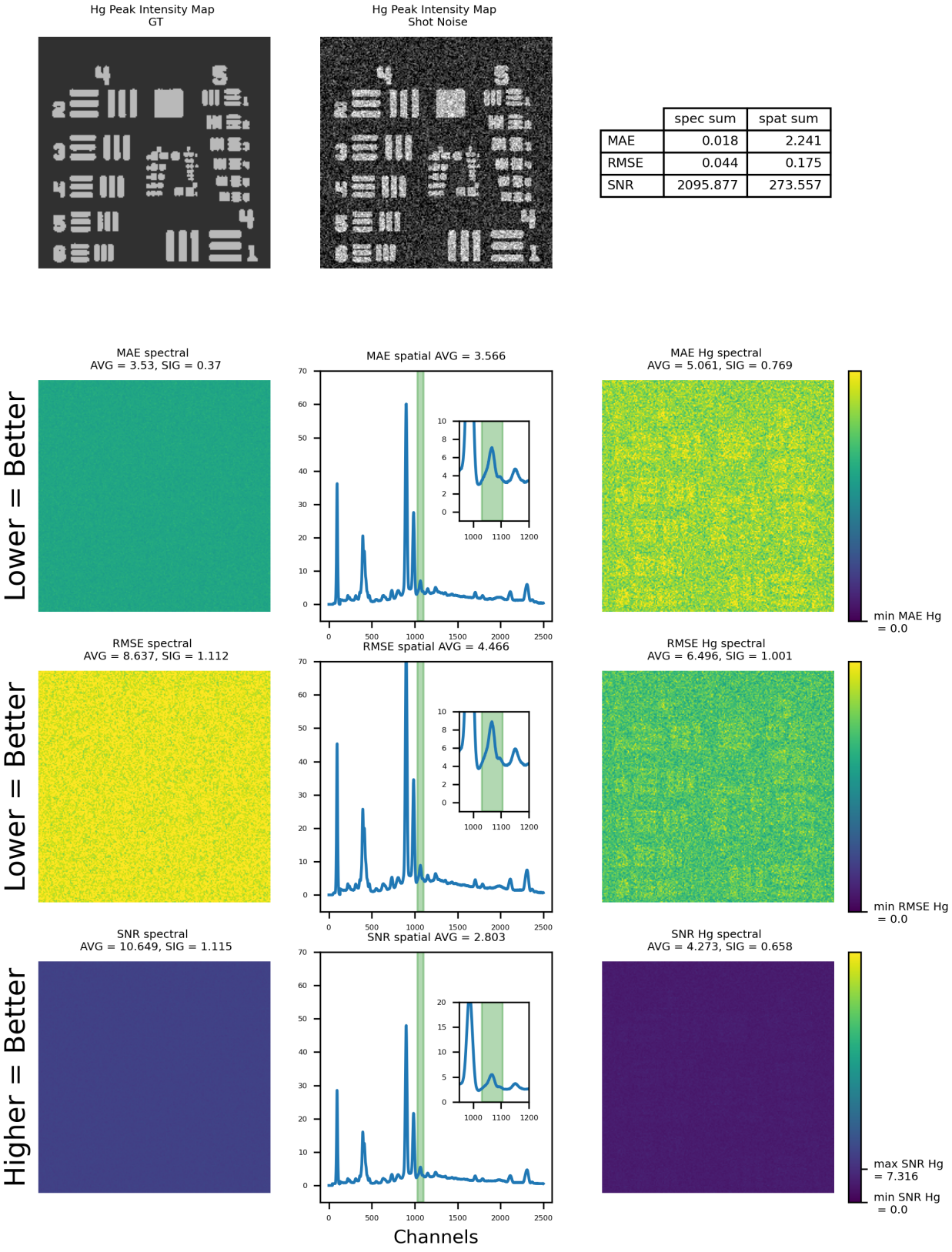
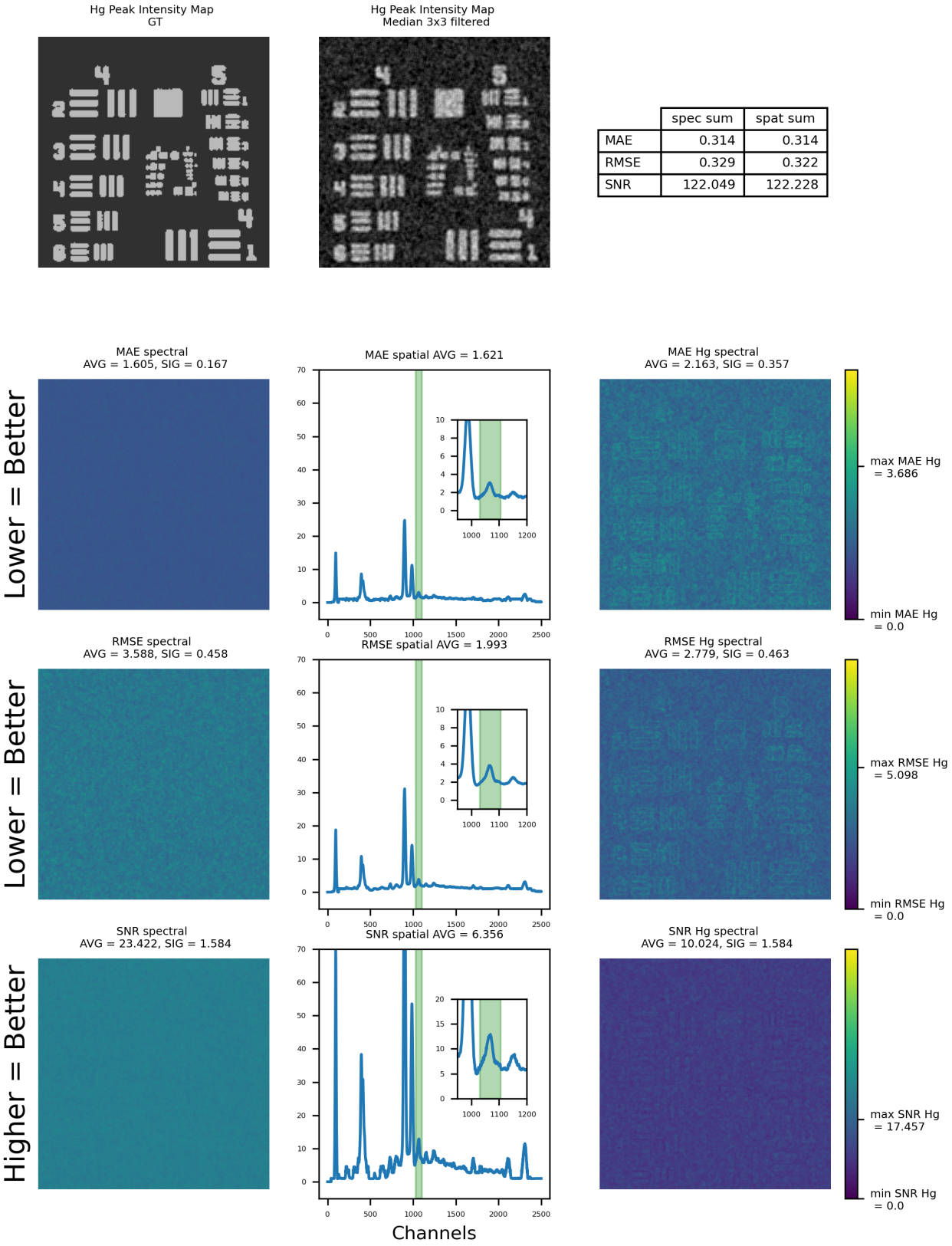


Figure 71: Performance Quality of Shot Noise-Polluted Data with a Dwell Time of 0.5 s.

Median 3x3 filtered data with a dwell time of 0.5 sec.



	spec sum	spat sum
MAE	0.314	0.314
RMSE	0.329	0.322
SNR	122.049	122.228

Figure 72: Performance Quality of 3x3 Median Filtered Data with a Dwell Time of 0.5 s.

HAAR LEV3 s3 filtered data with a dwell time of 0.5 sec.

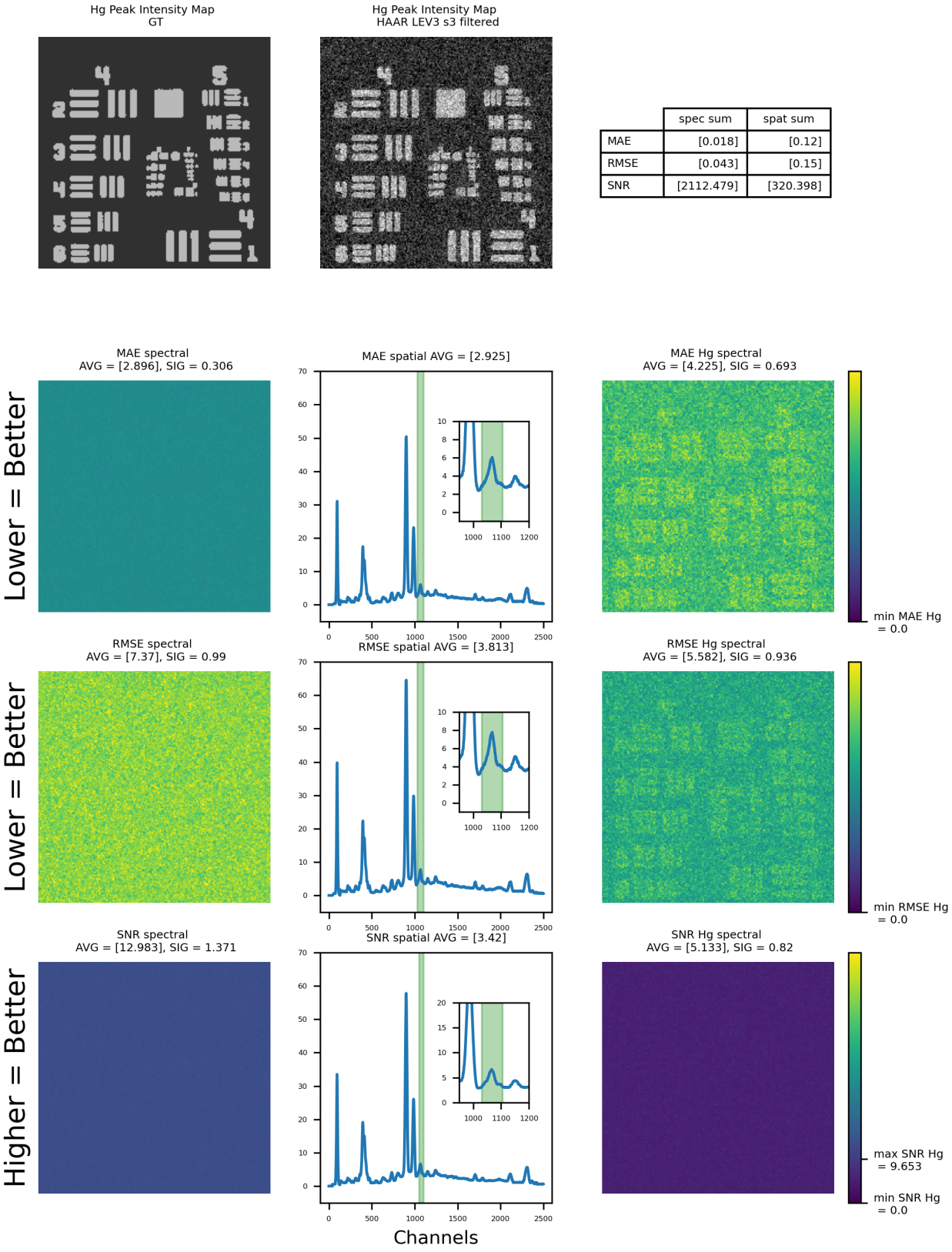
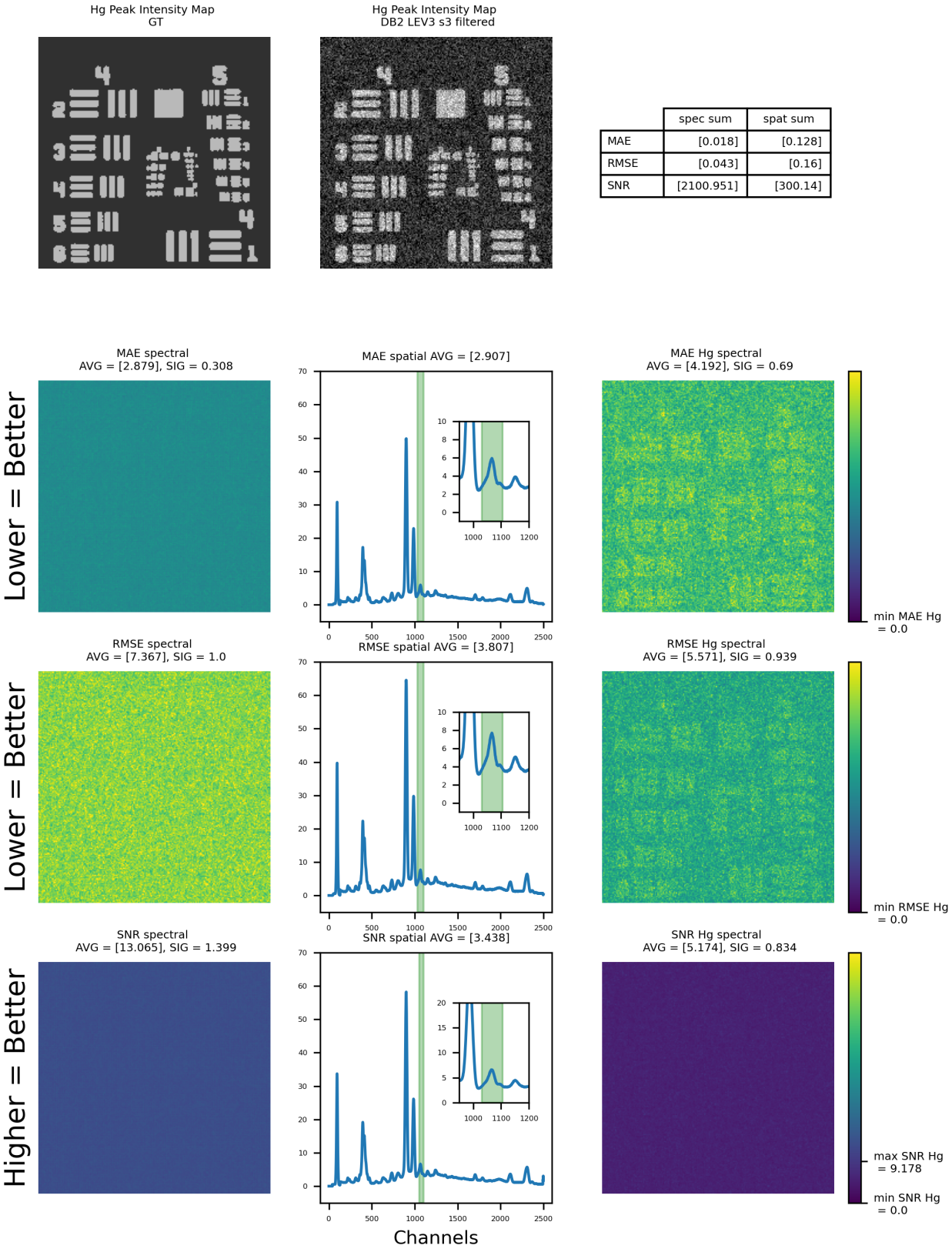


Figure 73: Performance Quality of haar Lev 3 s3 Filtered Data with a Dwell Time of 0.5 s.

DB2 LEV3 s3 filtered data with a dwell time of 0.5 sec.



	spec sum	spat sum
MAE	[0.018]	[0.128]
RMSE	[0.043]	[0.16]
SNR	[2100.951]	[300.14]

Figure 74: Performance Quality of db2 Lev 3 s3 Filtered Data with a Dwell Time of 0.5 s.

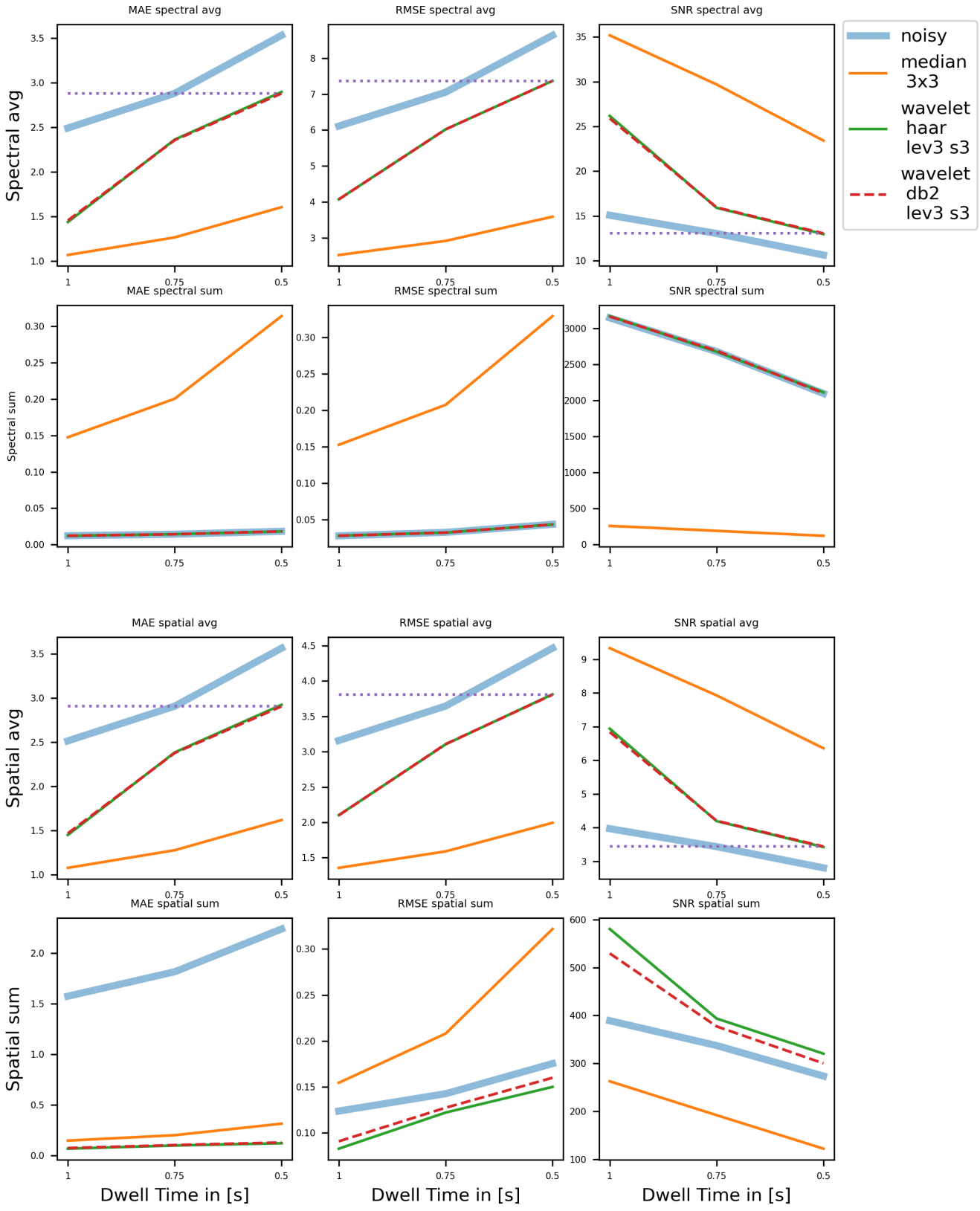


Figure 75: Performance quality of noisy data, median 3x3 filtered data, haar lev3 s3 filtered data and db2 Lev 3 s3 filtered data for dwell times 1, 0.75, and 0.5 sec. The intersection of the purple dotted line with the graph of the noisy data shows at what dwell time equal performance quality is obtained between the unfiltered data and the wavelet filtered data that was obtained with a dwell time of 0.5 sec.

5.4.2 Discussion Testing Round 3

In Figure 62, Figure 63, Figure 64, Figure 65, and Figure 66 it could be seen that the average performance quality of the median filters was better than that of the wavelets, however, their sum performance quality was significantly worse. Even worse than the sum of the raw data. When plotting the sum spectra in Figure 59, Figure 60, and Figure 61 of the filtered data it could indeed be seen that the median filtered data did not follow the ground truth, while the wavelet filtered data did. This probably has to do with the median filter altering the signal through over-averaging.

With more noise, median filtering still had a higher average performance quality than wavelet filtering. Similarly, the wavelets still had a higher sum performance quality. When regarding the sum spectra in Figure 59, Figure 60, and Figure 61, it was visible that with higher noise levels, median filtering becomes less reliable than wavelets in maintaining the original signal. The latter was reaffirmed by regarding Figure 75, where the spectral sum performance quality is nearly identical for the raw data and the wavelet filtered data.

In other words, similar to testing round 1 and 2, it could be seen that averaging, this time by median filters, resulted in good average performance quality. Yet, when summating the data, it becomes clear that wavelet denoising is a method that is better suited to stay true to the original signal.

Moreover, in Figure 75, it also became clear that the average performance quality of filtered data obtained with a dwell time of 0.5 seconds was approximately equal to the average performance quality of unfiltered data obtained with a dwell time of 0.75 seconds. Hence, the wavelet filter achieves the goal of lowering the necessary dwell time.

5.5 Real-world Case Study

Three of the filters that were compared in the third testing round are used on an actual MA-XRF dataset. We apply the haar lev3 s3, the db2 lev3 s3, and the 3×3 median filter to the daguerreotype shown in Figure 23. The daguerreotype has been scanned both with a dwell time of 30 ms and a dwell time of 100 ms. The unfiltered Hg L3 peak intensity maps are shown in Figure 76.

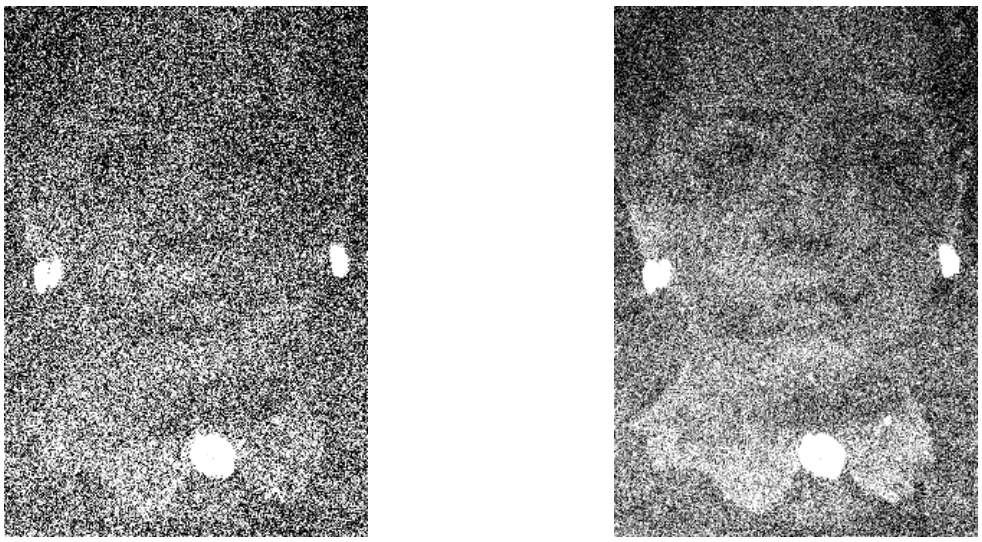


Figure 76: Unfiltered Hg L3 peak intensity maps. Left: scan with 30 ms dwell time, Right: scan with 100 ms dwell time

As expected, the scan of 30 ms looks grainier than the scan of 100 ms.

In Figure 77 and Figure 78, the data, which was obtained with a dwell time of 30 ms, after filtering is visualised. When comparing the median filtered to the wavelet-filtered data, it can be seen that the wavelet-filtered data retains the resolution of the data better. Comparing the two wavelet filters, it can be seen that the haar filtered peak intensity map is less smooth than the db2 map.

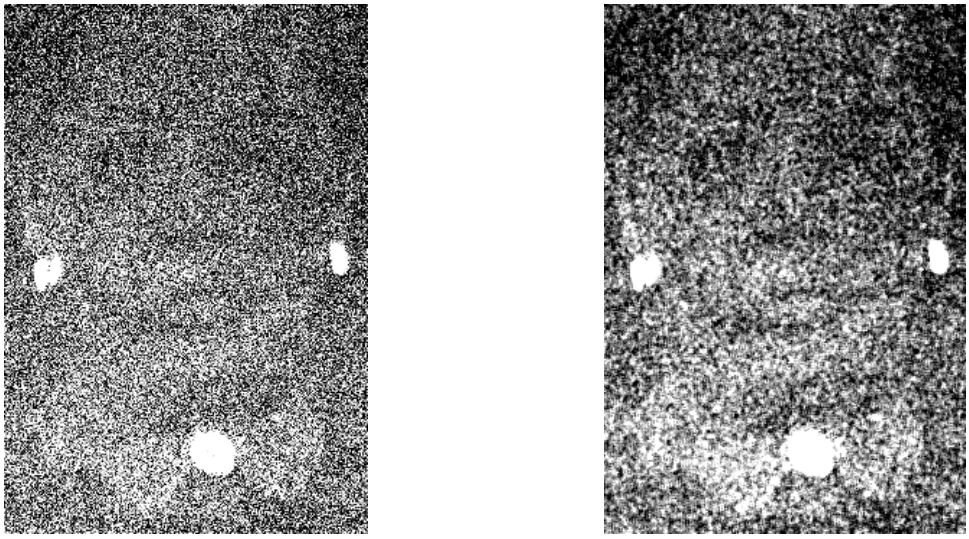


Figure 77: Unfiltered Hg L3 peak intensity maps (data obtained with 30 ms dwell time). Left: raw data, Right: median filtered data

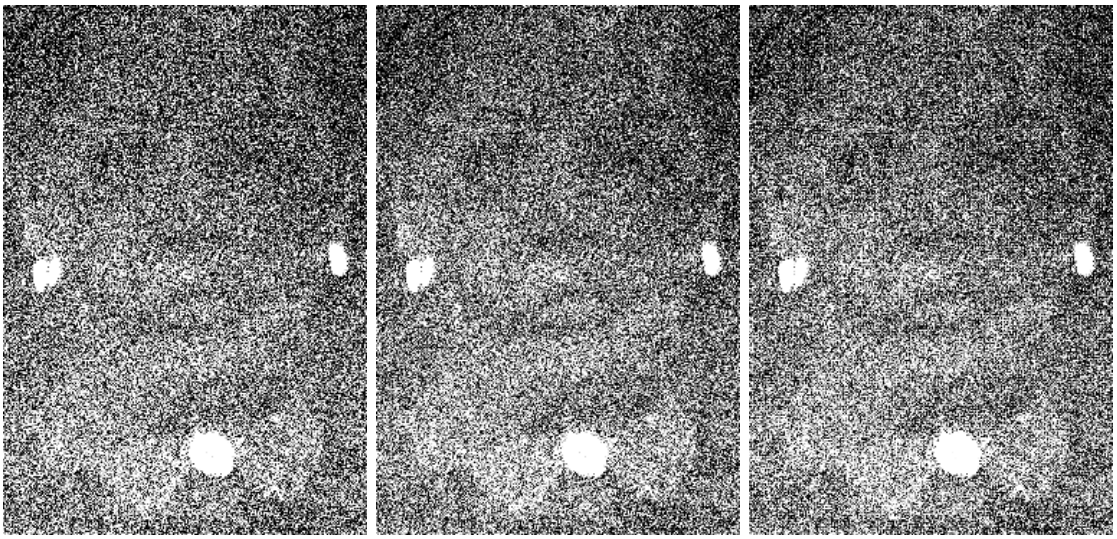


Figure 78: Filtered Hg L3 peak intensity maps (data obtained with 30 ms dwell time). Left: raw data, Middle: haar lev3 s3 filtered, Right: db2 lev3 s3 filtered

5.6 General Discussion and Future Work

Although the performed tests resulted in a well-considered choice of wavelet design parameters, still, care must be taken for a lot of the results are dataset dependent and cannot be assumed definitive for MA-XRF data in general.

On the methodology. The numerical results obtained within this thesis were not self-evident which has partly to do with the methodology. The synthetic dataset used for testing the filters is binary and has large patches of similar elemental distributions. Moreover, the difference between the photon counts of mercury-rich and mercury-poor areas was minor. Therefore, the numerical performance quality indicators are not able to properly capture and consequently put a penalty on over-smoothing,

On the Wavelet Filter bank Design. When it comes to the optimal support size or amount of decomposition levels, different values could be optimal based on the step size used for obtaining the data and the density of singularities within the data. There is also a trade-off between high level-decomposition and low level-decomposition. Too low decomposition levels will capture noise in the approximation coefficients while too high decomposition levels will capture signal in the detail coefficients.

In a similar manner, it could be seen how well the haar wavelet performed with respect to the other wavelet families. This probably had to do with the binary nature of the spatial pattern in the synthetic dataset that could be captured perfectly using flat haar wavelets. When looking at Figure 78, for the real-world dataset the db2 filter seems to get better results, visually at least. Also the optimal mother wavelet is not universal for every MA-XRF dataset.

We do hypothesize that the amount of vanishing moments that appeared to be optimal in this research is generally speaking a good fit for MA-XRF datasets, since the MA-XRF datasets in need of denoising are generally irregular and, hence, will need few vanishing moments. Besides, in section 5.1.2 it could be seen that the amount of vanishing points was more important than the support size (coif1 had good performance as well). So even for changes in the above-mentioned density of singularities and step size, the amount of vanishing moments should be small.

On the Thresholding Design. The fact that the universal thresholding method has been adopted allows for further improvement of the thresholding algorithm. The Poisson nature of the data and the need for spectral dimension awareness have been addressed by using neighbouring channels. However, there are most likely more advanced approaches possible.

Whether the number of neighbouring channels that were rendered optimal for this synthetic dataset is also optimal for other MA-XRF datasets is not likely. It strongly depends on the width of the elemental peaks, more specifically, the elemental peaks of interest. This also means that it is very important to take into account that these tests have been performed with data that was normalized to a zero peak intensity of 10,000. Moreover, the optimal number of neighbouring

channels depends on the amount of shot noise as well. Similar to the number of applied decomposition levels, there is a trade-off between high and low s. Too high values for s do not remove the noise and too low values for s over-average the data.

On the Co-Dependance of Decomposition Levels and Neighbouring Channels. In the second testing round, it was found that the optimal number of neighbouring channels depended on the decomposition level. Low levels of decomposition required low s for optimal performance and high levels of decomposition required high s for optimal performance. The optimal combination of the two is probably data-dependent and careful consideration is necessary for achieving denoising while abstaining from averaging.

On the Use of Wavelet Denoising for MA-XRF Data, Outlook. At the moment, not a lot can be said about the general applicability of the wavelet filter to filter MA-XRF data, to this end further testing is needed. Still, the wavelet denoising algorithm gives promising results that could be investigated further in future research.

One of the promising results could be seen in section 5.4, where the sum spectrum of the wavelet filtered data was equal to the sum spectrum of the unfiltered data. Hence, the signal is maintained, contrary to median filters, where the sum spectrum is significantly altered. So, despite the better performance of median filtering on average pixel and channel basis, the wavelet filter is better at preserving the signal while filtering.

For future research, I would recommend doing more experiments. Experiments in which the objects of research have different elemental compositions and various singularity densities. Moreover, experiments where the settings of the MA-XRF scanner differ in stepsize and dwell time. Furthermore, besides applying the method of this research to various datasets, it would also be interesting to experiment with different thresholding methods. For example, expanding upon the idea mentioned by Louisier *et al.*, who proposed linking detail coefficients to approximation coefficients in order to determine a better threshold for shot noise-polluted data,[57], by circumventing the need of prior information.

6 Conclusion

This research aimed to address the problem of long scanning times for MA-XRF. The scanning time could be reduced by lowering the dwell time, however, this would result in a low SNR. Therefore, it was proposed to increase the SNR post-

measurement by denoising. More specifically, this research sought to increase the spectral SNR using local spectral and spatial relations within the dataset.

By carefully evaluating the MA-XRF data characteristics, it was concluded that wavelet filter bank denoising would be a suitable method to denoise MA-XRF data. The four main data characteristics that were taken into account were the heterogeneous signal structure, the fact that the spatial signal has a localized nature, the spatial and spectral relation within the hyperspectral datacube, and the main source of noise in case of low dwell times: photon shot noise.

The wavelet filter bank denoising method does not require *a priori* information in order to filter the data, hence, the difficulty of the fact that every dataset is unique is overcome. Besides that, wavelets are able to capture the local and discontinuous signal characteristics of MA-XRF data: wavelets capture high-frequency signal features, while also achieving to denoise the data without losing these features. Moreover, spatial wavelet filtering succeeds in keeping the photon counts per channel intact through the perfect reconstruction principle.

Most thresholding methods used in wavelet filter bank denoising assume Gaussian noise and the ones that do take into account shot noise have properties that counteract one of the other three demands for a MA-XRF denoising algorithm. Therefore, a novel procedure was proposed in this work where use was made of neighbouring channels to determine a proper threshold value. Here, the effects of shot noise were counteracted by determining the standard deviation on more counts, Herewith lowering the standard deviation of the detail coefficients on a per-channel basis, allowing the threshold to be determined based on the measured average photon intensity fluctuations of the channel and its neighbouring channels.

There are different design parameters of the wavelet filter that could be optimized. In this research, universal thresholding as a threshold method has been adopted. However, we did optimize the number of neighbouring channels to determine the standard deviation of the detail coefficients used to calculate the universal threshold. Furthermore, we optimized for the type of mother wavelet and the number of decomposition levels.

The wavelet filters were tested on a synthetic MA-XRF dataset where mercury-rich and mercury-poor pixel spectra were distributed according to a 1951 USAF resolution test chart, using MAE, RMSE and SNR as performance indicators. Eventually, a mother wavelet having little vanishing moments was shown to be optimal for MA-XRF data, at least in the examined case studies. To this end the haar, db2, sym2, and coif1 wavelet can be used. Less evident is the optimal choice for the number of decomposition levels and neighbouring channels. It was, moreover found that the two are co-dependent. The higher the decomposition level, the higher the number of neighbouring channels and vice versa. The reason for this is that both a high decomposition level and a low number of neighbouring

channels result in over-smoothing, while a low number of decomposition levels and a high number of neighbouring channels result in noise not being removed. Hence, they need to be in balance. A clear optimal combination of the two could not readily be determined based on the performance quality indicators. This has most likely to do with the binary nature of the synthetic dataset, having large patches of similar elemental distributions, not penalizing over-smoothing. Moreover, the difference between the photon counts of mercury-rich and mercury-poor areas was minor. Still, by combining visual analyses with performance quality indicators determined on summated areas within the synthetic dataset, a suitable combination was found of 3 decomposition levels with $s = 3$ neighbouring channels. Applying the haar and db2 wavelet with a 3-level decomposition and $s = 3$ both resulted in data simulated to have a dwell time of 0.5 seconds to have approximately equal SNR to unfiltered data simulated to have a dwell time of 0.75 seconds. Herewith the wavelet filters adhere to the objective of lowering the dwell time.

For a generally applicable design of a wavelet filter for MA-XRF data, further testing is needed. It could be seen, for example, in the real-world daguerreotype case study that the spatial shape of the signal influenced the optimal mother wavelet. Furthermore, the singularity density and the stepsize influence the optimal support size of the wavelet and the optimal number of decomposition levels of the wavelet filter bank.

References

[1] E. Mansfield, *Art History and Its Institutions: Foundations of a discipline*. John Wiley Sons, LTD, 2000.

[2] M. Faries, "Reshaping the field: The contribution of technical studies," in *Early Netherlandish Painting at the Crossroads: A Critical Look at Current Methodologies*, M. W. Ainsworth, Ed. The Metropolitan Museum of Art, 2001, pp. 70–105.

[3] P. Romano and K. Janssens, "Preface to the special issue on: Ma-xrf "developments and applications of macro-xrf in conservation, art, and archeology" (trieste, italy, 24 and 25 september 2017)," *X-Ray Spectrometry*, vol. 48, pp. 249–250, 2019.

[4] M. Alfeld, "Ma-xrf for historical paintings: State of the art and perspective," *Microscopy and Microanalysis*, vol. 26, pp. 72–75, 2020. [Online]. Available: https://doi.org/10.1017/S1431927620013288

[5] M. Alfeld, K. Janssens, J. Dik, W. De Nolf, and G. Van der Snickt, "Optimization of mobile scanning macro-xrf systems for the in situ investigation of historical paintings †," *Journal of Analytical Atomic Spectrometry*, vol. 26, pp. 899–909, 2011. [Online]. Available: https://doi.org/10.1039/c0ja00257g

[6] K. Janssens, "Axes and the ghent altarpiece," https://www.uantwerpen.be/en/projects/axes-ghent-altarpiece/, accessed: 2022-06-04.

[7] H. Chopp, A. McGeachy, M. Alfeld, O. Cossairt, M. Walton, and A. Katsaggelos, "Denoising fast x-ray fluorescence raster scans of paintings," pre-print, 2022.

[8] G. Van der Snickt, H. Dubois, J. Sanyova, S. Legrand, A. Coudray, C. Glaude, M. Postec, P. Van Espen, and K. Janssens, "Large-area elemental imaging reveals van eyck s original paint layers on the ghent altarpiece (1432), rescoping its conservation treatment," *Angewandte Chemie International Edition*, vol. 56, pp. 4797–4801, 2017. [Online]. Available: https://doi.org/10.1002/anie.201700707

[9] H. Chopp, A. McGeachy, M. Alfeld, O. Cossairt, M. Walton, and A. Katsaggelos, "Image processing perspectives of x-ray fluorescence data in cultural heritage sciences," *IEEE BITS the Information Theory Magazine*, pp. 1–12, 2022.

[10] J. Oppenheimer, "On the quantum theory of the capture of electrons," *Physical Review*, vol. 31, pp. 349–356, 1928.

[11] "Handheld xrf: How it works," https://www.bruker.com/en/products-and-solutions/elemental-analyzers/handheld-xrf-spectrometers/how-xrf-works.html, accessed: 2022-05-15.

[12] O. Instruments, *Silicon Drift Detectors Explained*. Oxford Instruments, 2012.

[13] K. Laclavetine, D. Giovannacci, M. Radepont, A. Michelin, A. Tournié, O. Belhadj, and W. Andraud, C.and Nowik, "Macro x-ray fluorescence (ma-xrf) scanning, multi-and hyperspectral imaging study of multiple layers of paintings on paneled vault in the church of le quillio (france)," *X-Ray Spectrometry*, 2020. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03015532

[14] S. Yan, H. Verinaz-Jadan, J. Huang, N. Daly, C. Higgitt, and P. Dragotti, "Super-resolution for macro x-ray fluorescence data collected from old master paintings," *ResearchGate*, pre-print, 2022.

[15] "Xrf research inc." http://www.xrfresearch.com, accessed: 2022-05-17.

[16] M. Alfeld, J. V. Pedroso, M. Van Eikema Hommes, G. Van Der Snickt, G. Tauber, J. Blaas, M. Haschke, K. Erler, J. Dik, and K. Janssens, "A mobile instrument for in situ scanning macro-xrf investigation of historical paintings," *Journal of Analytical Atomic Spectrometry*, vol. 28, pp. 760–767, 2013. [Online]. Available: https://doi.org/10.1039/c3ja30341a

[17] R. Grieken van and A. Markowicz, *Handbook of X-Ray Spectrometry*. Marcel Dekker, 2002.

[18] M. Alfeld, G. Van der Snickt, F. Vanmeert, K. Janssens, J. Dik, K. Appel, L. Van der Loeff, M. Chavannes, T. Meedendorp, and A. Hendriks, "Scanning xrf investigation of a flower still life and its underlying composition from the collection of the kröller–müller museum," *Applied Physics A: Materials Science Processing*, vol. 111, pp. 165–175, 2013.

[19] "M6 jetstream," https://www.bruker.com/en/products-and-solutions/elemental-analyzers/micro-xrf-spectrometers/m6-jetstream.html, accessed: 2022-05-15.

[20] A. Martins, J. Coddington, G. Van der Snickt, B. Driel, C. McGlinchey, D. Dahlberg, K. Janssens, and J. Dik, "Jackson pollock's number 1a, 1948: A non-invasive study using macro-x-ray fluorescence mapping

(ma-xrf) and multivariate curve resolution-alternating least squares (mcr-als) analysis," *Heritage Science*, vol. 4, pp. 1–13, 2016. [Online]. Available: https://doi.org/10.1186/s40494-016-0105-2

[21] A. van Loon, P. Noble, A. Krekeler, G. Van der Snickt, K. Janssens, Y. Abe, I. Nakai, and J. Dik, "Artificial orpiment, a new pigment in rembrandt's palette," *Heritage Science*, vol. 5, pp. 1–13, 2017. [Online]. Available: https://doi.org/10.1186/s40494-017-0138-1

[22] A. Martins, C. Albertson, C. McGlinchey, and J. Dik, "Piet mondrian's broadway boogie woogie: Non invasive analysis using macro x-ray fluorescence mapping (ma-xrf) and multivariate curve resolution-alternating least square (mcr-als)," *Heritage Science*, vol. 4, pp. 1–16, 2016. [Online]. Available: https://doi.org/10.1186/s40494-016-0091-4

[23] M. Alfeld, V. Gonzalez, and A. van Loon, "Data intrinsic correction for working distance variations in ma-xrf of historical paintings based on the ar signal," *X-Ray Spectrometry*, vol. 50, pp. 351–357, 2021.

[24] M. Alfeld and K. Janssens, "Strategies for processing mega-pixel x-ray fluorescence hyperspectral data: A case study on a version of caravaggio's painting supper at emmaus," *Journal of Analytical Atomic Spectrometry*, vol. 30, pp. 777–789, 2015. [Online]. Available: https://doi.org/10.1039/c4ja00387j

[25] M. Alfeld, "Development of scanning macro-xrf for the investigation of historical paintings," 2013.

[26] M. Alfeld, M. Wahabzada, C. Bauckhage, K. Kersting, G. Wellenreuther, and G. Falkenberg, "Non-negative factor analysis supporting the interpretation of elemental distribution images acquired by xrf," *Journal of Physics: Conference Series*, vol. 499, pp. 1–10, 2014.

[27] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of Chemometrics*, vol. 11, pp. 393–401, 1997.

[28] C. Thurau, K. Kersting, M. Wahabzada, and C. Bauckhage, "Descriptive matrix factorization for sustainability adopting the principle of opposites," *Data Mining and Knowledge Discovery*, vol. 24, pp. 325–354, 2012.

[29] M. Alfeld, M. Wahabzada, C. Bauckhage, K. Kersting, G. van der Snickt, P. Noble, G. Wellenreuther, and G. Falkenberg, "Simplex volume maximization (sivm): A matrix factorization algorithm with non-negative constrains

and low computing demands for the interpretation of full spectral x-ray fluorescence imaging data," *Microchemical Journal*, vol. 132, pp. 179–184, 2017.

[30] M. Keenan and P. Kotula, "Optimal scaling of tof-sims spectrum-images prior to multivariate statistical analysis," *Applied Surface Science*, vol. 231-232, pp. 240–244, 2004.

[31] S. Kogou, L. Lee, G. Shahtahmassebi, and H. Liang, "A new approach to the interpretation of xrf spectral imaging data using neural networks," *X-Ray Spectrometry*, vol. 50, pp. 310–319, 2021.

[32] C. Ryan, E. Clayton, W. Griffin, S. Sie, and D. Cousens, "Snip, a statistics-sensitive background treatment for the quantitative analysis of pixe spectra in geoscience applications," *Nuclear Instruments and Methods in Physics Research*, vol. 34, pp. 396–402, 1988.

[33] B. Vekemans, K. Janssens, L. Vincze, and P. Adams, F. Van Espen, "Comparison of several background compensation methods useful for evaluation of energy-dispersive x-ray fluorescence spectra," *Spectrochimica Acta*, vol. 50B, pp. 149–169, 1994.

[34] K. Janssens, F. Adams, and A. Rindby, *Microscopic X-ray Fluorescence Analysis*. Routledge, 2002.

[35] P. M. Wróbel, P. Fraczek, and M. Lankosz, "New approaches for correction of interlayer absorption effects in x-ray fluorescence imaging of paintings," *Analytical Chemistry*, vol. 88, pp. 1661–1666, 2016.

[36] B. Vekemans, K. Janssens, L. Vincze, A. Aerts, F. Adams, and J. Hertogen, "Automated segmentation of μ-xrf image sets," *X-Ray Spectrometry*, vol. 26, pp. 333–346, 1997.

[37] J. Yang, Z. Zhang, and Q. Cheng, "Resolution enhancement in micro- xrf using image restoration techniques," *Journal of Analytical Atomic Spectrometry*, vol. 37, pp. 750–758, 2022.

[38] Q. Dai, E. Pouyet, O. Cossairt, M. Walton, and A. Katsaggelos, "Spatial-spectral representation for x-ray fluorescence image super-resolution," *IEEE Transactions on Computational Imaging*, vol. 3, pp. 432–444, 2017.

[39] Q. Dai, H. Chopp, E. Pouyet, O. Cossairt, M. Walton, and A. Katsaggelos, "Adaptive image sampling using deep learning and its applications on x-ray fluorescence image reconstruction," *IEEE Transactions on Multimedia*, vol. 22, pp. 77–87, 2018.

[40] G. Snickt van der, "James ensor's pigments studied by means of portable and synchrotron radiation-based x-ray techniques: Evolution, context and degradation," 2012.

[41] M. Vetterli, "Filter banks allowing perfect reconstruction," *Signal Processing*, vol. 10, pp. 219–244, 1986.

[42] D. Donoho and I. Johnston, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[43] M. Jansen, *Noise Reduction by Wavelet Tresholding*. Springer New York, NY, 2001.

[44] K. Stokfiszewski, K. Wieloch, and M. Yatsymirskyy, "An efficient implementation of one-dimensional discrete wavelet transform algorithms for gpu architectures," *The Journal of Supercomputing*, vol. 87, pp. 11 539–11 563, 2022.

[45] R. Lemaster, "The use of the wavelet transform to extract additional information on surface quality from optical profilometers," in *Advances in Wavelet Theory and their Applications in Engineering, Physics and Technology*, D. Baleanu, Ed. InTech, 2012, pp. 99–124.

[46] S. Mallat, *A Wavelet Tour of Signal Processing*. Elsevier Inc., 2009.

[47] C. Burrus, R. Gopinath, and H. Guo, *Wavelets and Wavelet Transforms, A Primer*. Prentice Hall, Inc., 1998.

[48] R. Ogden, *Essential Wavelets for Statistical Applications and Data Analysis*. Springer Science+Business Media New York, 1997.

[49] I. Daubechies, *Ten Lectures of Wavelets*. Springer-Verlag, 1992.

[50] J. Gomes and L. Velho, *From Fourier Analysis to Wavelets*. Springer, 2015.

[51] C. Akujuobi, *Wavelets and Wavelet Transform Systems and Their Applications, A Digital Signal Processing Approach*. Springer, 2022.

[52] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, pp. 1200–1224, 1995.

[53] ——, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comptes Rendus De L Academie Des Sciences Serie I-Mathematique*, vol. 319, pp. 1317–1322, 1994.

[54] S. Chang and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, pp. 1532–1546, 2000.

[55] C. He, J. Xing, J. Li, Q. Yang, and R. Wang, "A new wavelet threshold determination method considering interscale correlation in signal denoising," *Mathematical Problems in Engineering*, 2014.

[56] M. Srivastava, C. Anderson, and J. Freed, "A new wavelet denoising method for selecting decomposition levels and noise thresholds," *IEEE Access*, vol. 4, pp. 3862–3877, 2016.

[57] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of poisson-corrupted images," *Signal Processing*, vol. 90, pp. 415–427, 2010.

[58] Y. Altay and A. Kremlev, "Signal-to-noise ratio and mean square error improving algorithms based on newton filters for measurement ecg data processing," *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering*, pp. 1590–1595, 2021.

[59] D. Balbas, B. Cattaneo, A. Cagnini, P. Belluzzo, S. Rossi, R. Fontana, and J. Striova, "The degradation of daguerreotypes and the relationship with their multi-material structure: A multimodal investigation," *Sensors*, vol. 23, 2023.

[60] M. Kozachuk and J. McElhone, "Applying nanoscience to daguerreotypes: Understanding and preserving the first form of the photograph," *National Gallery of Canada Review*, vol. 9, pp. 36–49, 2018.

[61] J. Davis and E. Vicenzi, "Optimizing compositional images of daguerreotype photographs using post processing methods," *Heritage Science*, vol. 4, 2016.