Processing and indexing of electron backscatter patterns using open-source software

IOP Conf. Series: Materials Science and Engineering **891** (2020) 012002 doi:10.1088/1757-899X/891/1/012002

Processing and indexing of electron backscatter patterns using open-source software

H W Ånes[1], J Hjelen[1], B E Sørensen[2], A T J van Helvoort[3] and K Marthinsen[1]

[1] Norwegian University of Science and Technology, Department of Materials Science and Engineering, 7491 Trondheim, Norway
[2] Norwegian University of Science and Technology, Department of Geoscience and Petroleum, 7491 Trondheim, Norway
[3] Norwegian University of Science and Technology, Department of Physics, 7491 Trondheim, Norway

E-mail: hakon.w.anes@ntnu.no

Abstract. A new method to increase the signal-to-noise ratio *S/N* of electron backscatter patterns (EBSPs) based upon principal component analysis (PCA) is presented. The PCA denoising method is applied to ten scans of EBSPs from the same region of interest of a recrystallised nickel sample acquired with a decreasing *S/N*, achieved by reducing the exposure time while increasing the camera gain accordingly. That PCA denoising increases *S/N* in EBSPs is demonstrated by comparing indexing success rates after both Hough and dictionary indexing (HI and DI) of the Ni patterns having undergone one of four processing routes: i) standard static and dynamic background corrections (standard corrections), ii) standard corrections and pattern averaging with the four closest neighbours, iii) standard corrections and PCA denoising, and iv) standard corrections and pattern averaging followed by PCA denoising. Both pattern averaging and PCA denoising increases the indexing success rates for both indexing approaches for the studied Ni scans, with the former processing route providing the better success rates. The best success rates are obtained after pattern averaging followed by PCA denoising. The potential of PCA denoising to reveal additional pattern details compared to standard corrections and pattern averaging is demonstrated in a pattern from an orthoclase ($KAlSi_3O_8$) grain in a geological sample. Software code, and the Ni data sets, are released alongside this article as part of KIKUCHIPY, an open-source software package dedicated to processing and analysis of EBSPs.

1. Introduction

Electron backscatter diffraction (EBSD) in the scanning electron microscope (SEM) is a technique to characterise crystallographic features [1]. Data collection and indexing is reliable and automated through commercial systems. However, in recent years, open-source software packages have emerged as alternatives that are free to use, reveal all parts of their algorithms, and can facilitate swifter implementation of new concepts. An illustrative example is contributions to the indexing of electron backscatter patterns (EBSPs), such as dictionary indexing (DI) [2] available in the EMSOFT package, and the ASTROEBSD package [3]. While the latter package and commercial systems in general rely on extraction of Kikuchi bands in patterns using Hough-transform based indexing (HI), the DI approach compares dynamically simulated patterns to the full experimental patterns. DI can therefore index

patterns in which extraction of bands is difficult due to a lower signal-to-noise ratio *S/N* [4, 5], which can result from highly deformed or fine-grained materials, or high EBSD camera gain and short exposure times.

An area explored to a lesser extent in open-source software for EBSD analysis is to increase *S/N* before indexing, with pattern averaging within a kernel [5, 6] a recent example. Another possibility is to use multivariate statistical analysis (MSA), which aims to help extract the most useful information within the data. Previous applications of MSA to EBSPs have focussed on reducing the number of patterns to a set of significant component patterns [7-9], so-called dimensionality reduction, and then index only these patterns. However, one can also reconstruct the full set of EBSPs from the component patterns considered to contain useful signal while discarding the component patterns considered to contain mostly noise, effectively increasing *S/N*.

Here, we present the use of dimensionality reduction to increase *S/N* in EBSPs using principal component analysis (PCA) [10], so-called PCA denoising. PCA and other MSA methods are available via the open-source, multi-dimensional data analysis package HYPERSPY [11], which itself depends upon SCIKIT-LEARN [12]. The PCA denoising approach is outlined. Indexing results from HI and DI of patterns from ten EBSD scans from the same region of interest (ROI) of a recrystallised nickel (Ni) sample, acquired with a decreasing *S/N*, are compared. The patterns have undergone one of four processing routes: i) standard static and dynamic background corrections (standard corrections), ii) standard corrections and pattern averaging with the four closest neighbour patterns, as presented by Wright *et al.* [5], iii) standard corrections and PCA denoising, and iv) standard corrections and pattern averaging followed by PCA denoising. Also, details revealed in a pattern from a grain of orthoclase ($KAlSi_3O_8$) in a geological EBSD data set after either PCA denoising or pattern averaging are compared to details in a dynamically simulated pattern best matching the experimental pattern. The software code, and the Ni data sets, are released alongside this article as part of KIKUCHIPY [13], an open-source software package dedicated to processing and analysis of EBSPs.

2. Denoising of EBSPs with principal component analysis

Briefly, a map of EBSPs as shown in Fig. 1a is a set of n observations (here $n = n_x \times n_y$ patterns) of a number of s variables (here $s = s_x \times s_y$ detector pixel intensities). The goal of PCA is to decompose data in this format to a set of values of linearly uncorrelated, orthogonal variables called principal components (PCs), while retaining as much as possible of the variation in the data [10]. Before finding the PCs, the patterns are arranged into a matrix $\mathbf{X}$ of dimensions $(n \times s)$, as shown in Fig. 1b. Each pattern n_i in $\mathbf{X}$ is centred about their mean. A singular value decomposition (SVD) $\mathbf{X} = \mathbf{U\Sigma V}^T$ is performed where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\mathbf{\Sigma}$ is a diagonal matrix, shown in Fig. 1c, with respective dimensions $(n \times p)$, $(p \times p)$, and $(s \times p)$, with p as the rank of $\mathbf{X}$. The PCs in $\mathbf{V}^T$ is in our case a set of component factors p sorted in order of decreasing variance, describing the underlying intensity distributions on the detector. The product $\mathbf{U\Sigma}$, the component loadings p, is then the variation of their strength from one observation point to the next.

Ideally, the first component corresponds to the crystallographic feature most prominent in the data, for example the largest grain, the next corresponds to the second largest feature, and so on, until the later components at some point contain only noise. If this is the case we can increase *S/N* in the patterns by reconstructing our data matrix $\mathbf{X}_{\text{new}} = \mathbf{U\Sigma V}^T$, keeping only the $k \leq p$ components considered to contain useful signal, as shown in Fig. 1d.

3. Methods

3.1. Materials

A recrystallised, polycrystalline sample of pure Ni was selected to demonstrate improved indexing results of patterns with a lower *S/N*. As the final preparation steps the sample was electropolished with an A2 electrolyte followed by plasma cleaning. To highlight the additional pattern details revealed by

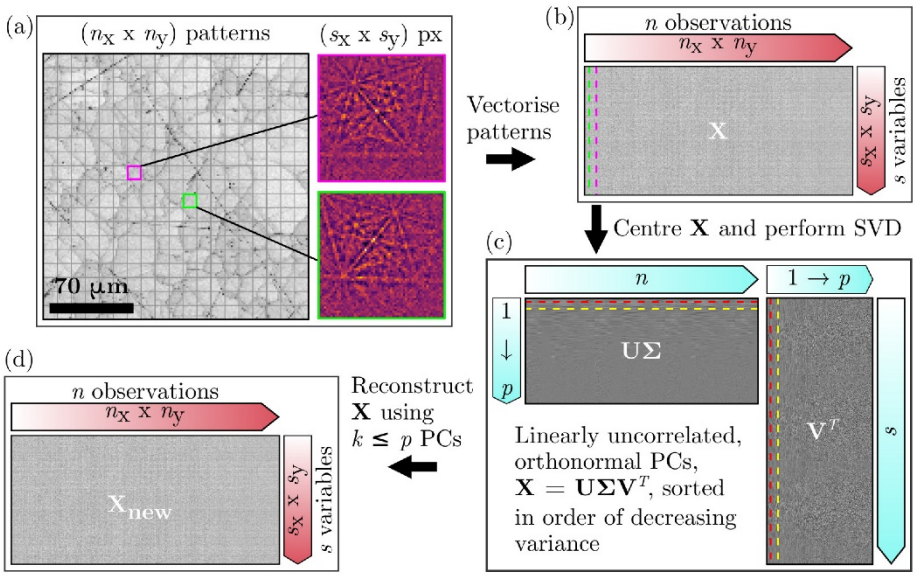

Figure 1. Processing steps to denoise (a, b) a data set of EBSPs **X** by first computing the (c) principal components p and then performing dimensionality reduction by reconstructing the data set $\mathbf{X}_{new}$ keeping only $k \times p$ components containing useful signal.

PCA denoising, a geological sample containing an orthoclase grain was studied. As the final preparation step this sample was polished with a suspension of 40 nm amorphous silica particles for 15 minutes and then rinsed in ethanol.

3.2. EBSD experiments

All EBSD data sets were obtained with a NORDIF UF-1100 camera on a Hitachi SU-6600 FEG SEM operated at 20 kV. The ten Ni data sets were collected consecutively from the same ROI with increasing camera gain and hence shorter exposure times listed in Table 1. 29,800 patterns with (60 × 60) pixels of 8 bit depth were collected for each scan in a (300 × 223.5) µm^2 square grid with a 1.5 µm step size. The geological data set comprise 94,248 patterns with (240 × 240) pixels of 8 bit depth collected in a (308 × 306) µm^2 square grid with a 5.0 µm step size and an 18.13 ms exposure time. Note that only patterns from a single grain in this data set are studied in this work.

Table 1. Scan number, camera gain, exposure time t and indices k of the principal components used in dimensionality reduction for the Ni data sets.

Scan	1	2	3	4	5	6	7	8	9	10
Gain [dB]	0	3	6	9	12	15	17	20	22	24
t [ms]	3.50	2.65	1.95	1.35	0.95	0.65	0.50	0.37	0.31	0.25
k	1-300	1-300	1-300	1-300	3-300	3-300	3-300	5-300	5,8-300	7-300

3.3. Pattern processing

All pattern processing was performed with KIKUCHIPY unless otherwise stated. The slowly varying diffuse background was removed from all patterns following standard background corrections. The correction procedure involves an initial subtraction of the raw pattern by a static background obtained by averaging multiple patterns collected over an area slightly larger than the ROI, followed by the subtraction by a dynamic (per pattern) background obtained by Gaussian blurring each pattern obtained after the first step.

Pattern averaging with the four closest neighbours was performed as indicated in Fig. 1 in Ref. [5]. Specifically, a (3 × 3) kernel, with all coefficients set to 1 except for 0 in the corners, was spatially correlated with the ($n_x \times n_y$) patterns, followed by normalisation of each pattern's intensity by dividing by the sum of kernel coefficients, which in this case was 5. Note that the implementation of neighbour pattern averaging in KIKUCHIPY allows for arbitrary 1D and 2D kernel sizes and coefficients, e.g., a Gaussian kernel.

The p first component patterns and loadings of each EBSD data set were found by performing PCA as explained above in section 2. After visual inspection, a number $k \leq p$ components, different for each data set, were chosen for dimensionality reduction. The components chosen for the Ni EBSD data are listed in Table 1, while up to 500 components were chosen for the geological EBSD data.

3.4. Indexing and indexing success rate

EDAX TSL 7.2 data collection was used for HI on the Ni patterns stretched to (80 × 80) pixels with a (9 × 9) convolution mask, a 2° θ step size, and searching for a maximum of 9 peaks, as was done in Ref. [5]. EMSOFT 4.2 was used for DI, with creation of dynamically simulated master patterns for both Ni [14] and orthoclase ($KAlSi_3O_8$) [15] with a 20 keV incident electron beam, a sample tilt of 70°, Monte Carlo energy bins from 10 - 20 kV with a 1 kV step, and to a maximum penetration depth of 100 nm with a 1 nm step. Simulated patterns were sampled from a uniform grid in orientation space with an average angular step of 1.4°, and the best matching orientation was further refined using a bound optimisation by quadratic approximation approach [16]. Pattern centres for both sets of EBSD data are given in Table 2. Note that the camera was binned from (480 × 480) px to (60 × 60) px and (240 × 240) px for the Ni and geological data sets, respectively. Before DI, both experimental and simulated patterns were processed in EMSOFT using a high-pass FFT filter with a maximum width of 0.125 and adaptive histogram equalisation with 4 and 5 regions for the Ni data and orthoclase pattern, respectively.

Table 2. Pattern centres in TSL (x^*, y^*, z^*) and EMSOFT 4.2 (x_{pc}, y_{pc}, L) conventions for the Ni and geological data sets.

Data	(x^*, y^*, z^*)	CCD pixel size [µm]	$(x_{\mathrm{pc}}, y_{\mathrm{pc}})$	L [µm]
Ni	(0.4210, 0.7794, 0.5049)	70.0	(-38.86, 132.65)	16 986.48
Orthoclase	(0.4033, 0.8739, 0.5115)	70.0	(-46.42, 179.47)	17 186.40

To assess the improvements to indexing results after PCA denoising and pattern averaging for the Ni data sets, each orientation in the scans was compared to orientations in the identical and adjacent points in a (3 × 3) kernel in scan 1, taken as the reference. The reference scans for the DI and HI data were the DI and HI results at zero camera gain and standard background corrections, respectively. If the orientation had a misorientation smaller than 5° to any of these points, it was considered to be successfully indexed. This calculation of the indexing success rate using a reference kernel (ISR^{RK}) [5], and the visualisation of orientation maps, were done with the open-source MATLAB texture toolbox MTEX [17] (note that a MATLAB license is required to run MTEX).

4. Results and discussion

EBSPs, acquired from the same ROI in Ni scan number 1, 6 and 9 after only standard background corrections (here called standard patterns) and after PCA denoising (here called denoised patterns) using the components indicated in Table 1, are shown in Fig. 2a. Also shown in the figure is a part of the orientation maps from scan 1 (left column), 6 (middle column) and 9 (right column) after Hough indexing of (b) standard patterns, (c) denoised patterns, (d) averaged patterns, and (e) averaged and denoised patterns, and (f) dictionary indexing of standard patterns, with their indexing success rates, ISR^{RK}, in the upper right corner. The signal-to-noise ratio in the original patterns decreases with increasing camera gain, however pattern averaging and PCA denoising retains some of the useful signal, as can be seen especially when comparing row (b) and (e) for scan 6 and 9. Among the scans and processing routes presented in Fig. 2, dictionary indexing of standard patterns provides the best results, except for HI of patterns in scan 9 after averaging and denoising, with respective ISR^{RK} of 69 % and 86.1 %.

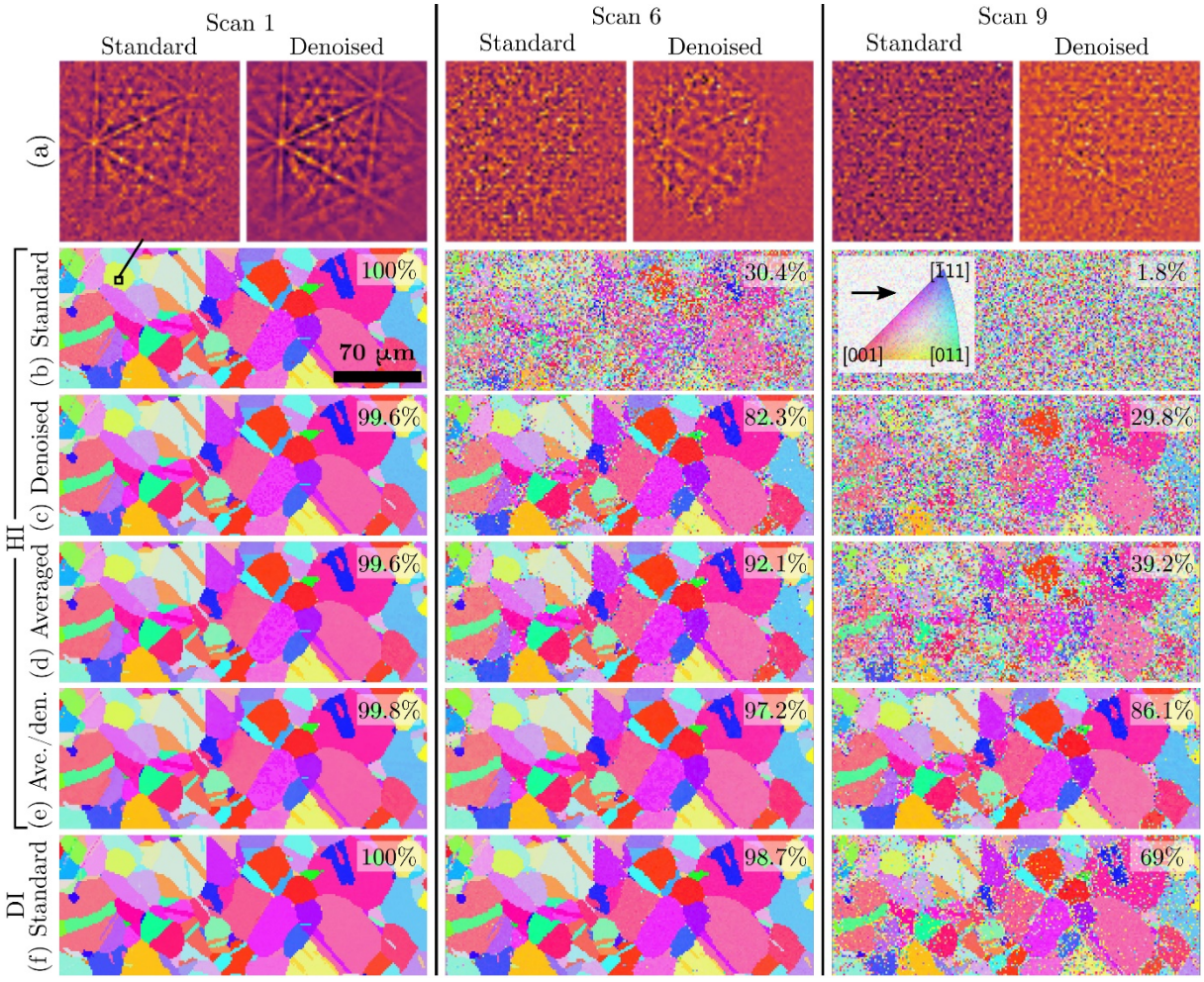

Figure 2. (a) Standard and denoised EBSPs from the same, indicated position in Ni scan number 1 (left), 6 (middle) and 9 (right) with orientation maps after HI of (b) standard patterns, (c) denoised patterns, (d) averaged patterns, and (e) averaged and denoised patterns, and (f) DI of standard patterns. The colour designates the crystal direction parallel to the right according to the colour key indicated. The ISR^{RK} is shown in the top right corner of each map.

The indexing success rates of all Ni scans after Hough and dictionary indexing of standard, averaged, denoised, and averaged and denoised patterns are shown in Fig. 3 as a function of camera gain. The DI ISR^{RK} of averaged and denoised patterns are not shown because a negligible increase in the success rates was observed compared to DI of averaged patterns. The HI and DI ISR^{RK} of standard patterns decreases from 100 % for scan 1 with zero camera gain, to 1 % and 46 % respectively for scan 10 with maximum gain. There are minor differences in the ISR^{RK} between the two indexing approaches until scan 5 where it drops to 65 % for HI, however, PCA denoising raises it to 94 %. PCA denoising also improves the DI ISR^{RK}, with the largest improvement, from 46 % to 61 %, for scan 10. Pattern averaging provides the better ISR^{RK} compared to PCA denoising, however, combining the two drastically improves the results obtained from HI. For scan 10, the HI ISR^{RK} jumps from a success rate of 17.5 % for averaged patterns to a success rate of 61.3 % for averaged patterns followed by PCA denoising. These results also show that HI might be sufficient in cases where patterns have a sufficient *S/N*. Furthermore, DI is highly robust towards noise, and should therefore be used when indexing patterns with a low *S/N* resulting in insufficient extraction of Kikuchi bands, in line with previous studies [4, 5]. This is best highlighted for scan 10, where almost all averaged patterns were successfully indexed with DI.

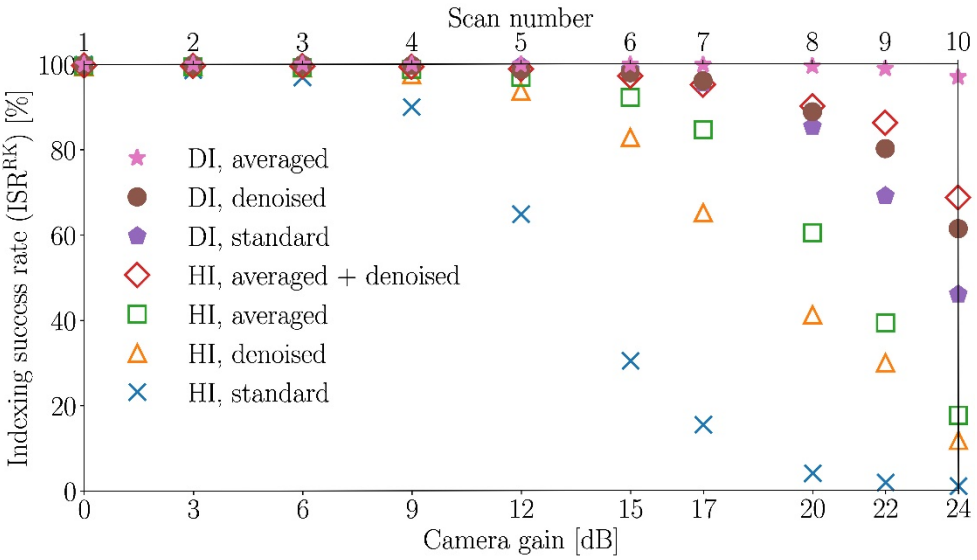

Figure 3. Comparison of the Hough and dictionary indexing success rates ISR^{RK} of standard, averaged, denoised, and averaged and denoised patterns, as a function of camera gain.

It is crucial to know how many principal components k to use for dimensionality reduction without loss of useful signal, as was discussed by Wilkinson *et al.* [8] and McAuliffe *et al.* [9]. The so-called scree plot of the explained variance as a function of the components sorted in order of decreasing variance is often used in multivariate statistical analysis to select this number [10]. Ideally, a significant change in the explained variance from a certain component to another can be identified, and thus the number of components to use. However, the explained variance decreases smoothly for the PCA results from the scans studied in this work, as was also observed by Wilkinson *et al.* [8]. To further complicate the selection, it is observed in this work that for the scans acquired with higher camera gain, notably scan 5 - 10, the first few components with the highest explained variance contained only noise and no discernible useful signal. These are the first few discarded components in Table 1. This demonstrates the need to inspect each principal component when selecting which components to use for dimensionality reduction for denoising of EBSPs.

To assess the number of components to use for dimensionality reduction for scan 9, the HI and DI ISR^{RK} after denoising with different number of components were calculated as shown in Fig. 4. 20 data sets were reconstructed with components in the range from 44 to 944 (from 50 to 1,000 minus the six discarded components listed in Table 1) in steps of 50. Interestingly, the trend in the ISR^{RK} is different for HI and DI. The HI ISR^{RK} is highest at 47 % when using only 44 components and decreases slowly with more components. The DI ISR^{RK} is lowest at 63 % when using 44 components and increases to 81 % with 244 components where it evens out. No significant decrease is observed with more components. An explanation for this trend in the DI ISR^{RK} is that the earlier components contain mostly signal from larger, similar crystallographic features, like grains, with the mid-range components up to about 244 containing a mix of signal from smaller grains and noise, until the later components contain only noise. An explanation for the decrease in the HI ISR^{RK} is that even though more components with signal from smaller grains are added to patterns from these grains, no indexing solution is found for these patterns. Instead, noise is added to patterns from larger grains, reducing the indexing success for these patterns. It is clear from Fig. 4 that most of the useful signal in scan 9 is found in the first 250 components, and thus the principal components up to $k = 250$, minus the first containing only noise, should be used for PCA denoising of this scan. Parts of the results from EMSOFT's DI is an indexing success rate, which is different from the ISR^{RK}, involving a computation of the smallest misorientation between the top matching orientation and the next nearest matches selected by the user. If the smallest misorientation is smaller than a threshold value the indexing is considered to be a success. This indexing success rate may be used to perform a search for the number of components to use for dimensionality reduction in cases where a reference scan is not available.

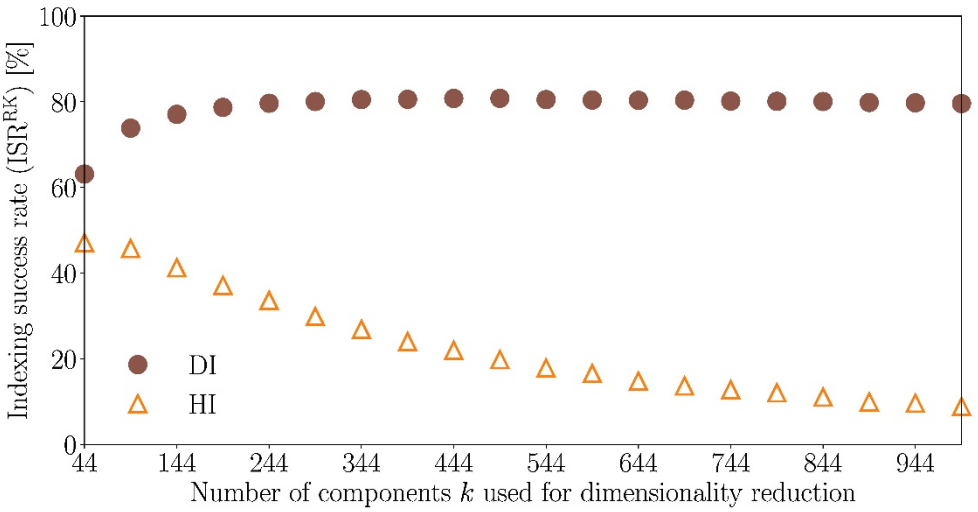

Figure 4. HI and DI ISR^{RK} as a function of number of principal components k used for dimensionality reduction in PCA denoising of scan 9.

The ability of PCA denoising to reveal fine details in EBSPs is studied in patterns from an orthoclase grain in a geological EBSD data set. A pattern from the grain after standard background corrections is shown in Fig. 5a. The average of 365 patterns containing Kikuchi bands from the grain is shown in (b), while the pattern from the same position as the one in (a) after PCA denoising is shown in (c). The best matching simulated pattern in (d) corresponds to the Euler angle triplet (261.1°, 36.0°, 272.3°). All patterns (a-d) underwent high-pass FFT filtering and adaptive histogram equalisation in EMSOFT as final processing steps. The dot product between the experimental and simulated patterns are shown in (a-c),

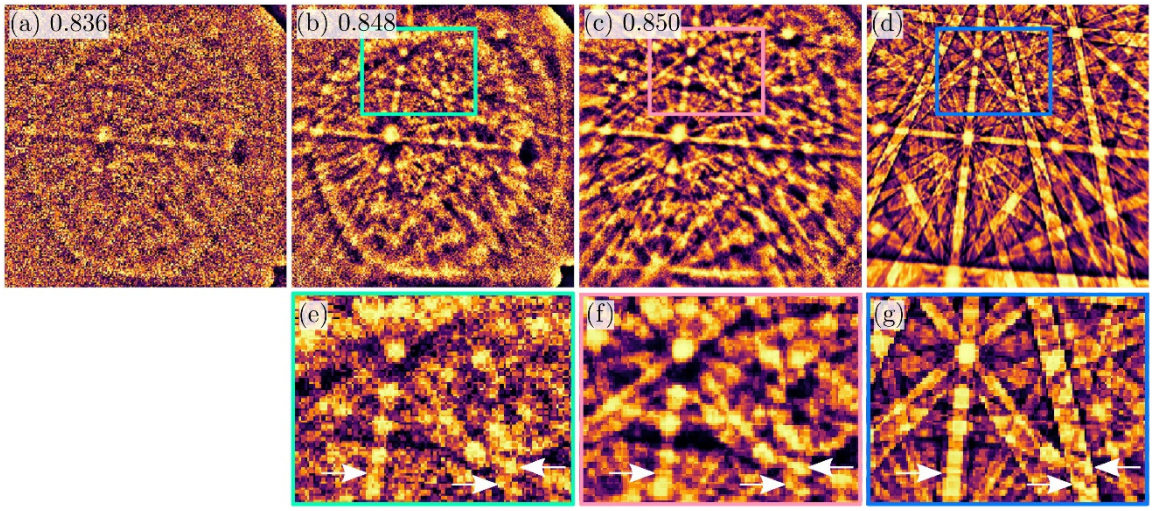

Figure 5. Comparison of an experimental pattern from an orthoclase grain after (a) standard background corrections, (b) averaging with all patterns within the grain, and (c) PCA denoising to (d) the simulated pattern with the highest dot product. The dot products are indicated, with the highest, 0.850, for the pattern in (c). Cut outs indicated in (b), (c), and (d) are magnified in (e), (f), and (g), respectively, and highlight features revealed by PCA denoising.

with the highest for the PCA denoised pattern. This and the fine details in the contrast from overlapping Kikuchi bands highlighted in the cut outs (e-g) demonstrate that PCA denoising in this example better harnesses the oversampling of the orthoclase grain than pattern averaging. Hansen *et al.* [18] found that Poisson-type noise, introduced into simulated patterns artificially and into experimental patterns by reducing the camera exposure time, can greatly inhibit the accuracy of geometrically necessary dislocation densities calculated from cross-correlation EBSD. Increasing *S/N* in patterns via PCA denoising might improve results from this type of analysis.

It should be noted that dictionary indexing is computationally intensive, mainly because each experimental pattern is compared to all patterns in the dictionary. For new users, the steps needed to create the dictionary of dynamically simulated EBSD patterns [19] might be challenging compared to performing Hough indexing via commercial software. However, it is our belief that these drawbacks are more than compensated for when considering the indexing results obtained after dictionary indexing of patterns of low *S/N* demonstrated in this work.

The patterns denoised in this work were obtained from recrystallised Ni and a large orthoclase grain. For the former, noise was introduced by reducing the exposure time and increasing the camera gain. A study of how PCA denoising can improve the quality of noisy patterns caused by high deformation and/or a fine-grained microstructure, related to the size of the interaction volume, is a natural continuation of this work. To improve upon the method of PCA denoising itself, a more robust selection of which principal components to keep for dimensionality reduction should be explored.

5. Conclusions

A method to increase the signal-to-noise ratio *S/N* in EBSPs based upon principal component analysis followed by dimensionality reduction, so-called PCA denoising, was presented. Applying PCA denoising to EBSD scans of Ni with a decreasing *S/N* demonstrated its ability to improve the indexing success rates after both Hough and dictionary indexing, effectively increasing the patterns' *S/N*. By first averaging each pattern with its four closest neighbours and subsequently applying PCA denoising, the *S/N* was increased further, and for these patterns the indexing success rates for Hough indexing was actually better than for dictionary indexing of patterns having undergone only standard background

corrections. That PCA denoising is able to reveal fine details in EBSPs not visible after standard background corrections or pattern averaging was also shown by visual comparison of a pattern after these three processing routes and a simulated pattern. The choice of principal components to keep for dimensionality reduction was discussed, highlighting the need for a careful selection based upon the *S/N* in EBSPs and the number of expected crystallographic features in a scan. Pattern processing before indexing was conducted with a new open-source software package for processing and analysis of EBSPs, named KIKUCHIPY. In releasing this software package and the Ni data sets, we hope to make a useful contribution to the growing suite of open-source software for processing and analysis of EBSPs.

6. Data statement

The Ni data sets can be downloaded from Zenodo (https://doi.org/10.5281/zenodo.3265037). An up-to-date copy of KIKUCHIPY is available from the GitHub repository (https://github.com/kikuchipy/kikuchipy).

Acknowledgments

HWÅ acknowledges the Norwegian University of Science and Technology (NTNU) for financial support through the NTNU Aluminium Product Innovation Centre (NAPIC). The authors thank the staff at the Electron Microscopy Lab at NTNU for maintaining the microscopy facilities. The authors are also grateful to the reviewer for suggesting we explore neighbour pattern averaging.

References

[1] Schwartz A J, Kumar M, Adams B L and Field D P 2009 *Electron backscatter diffraction in materials science. 2nd edition*. (New York, NY: Springer)
[2] Chen Y H, Park S U, Wei D, Newstadt G, Jackson M A, Simmons J P, De Graef M and Hero A O 2015 *Microsc. Microanal.* **21** 739-752
[3] Britton T B, Tong V, Hickey J, Foden A and Wilkinson A 2018 *J. Appl. Crystallography* **51** 1-10
[4] Singh S, Guo Y, Winiarski B, Burnett T L, Withers P J and De Graef M 2018 *Sci. Rep.* **8** 1-8
[5] Wright S I, Nowell M M, Lindeman S P, Camus P P, De Graef M and Jackson M A 2015 *Ultramicroscopy* **159** 81-94
[6] Brewick P T, Wright S and Rowenhorst D J 2019 *Ultramicroscopy* **200** 50-61
[7] Brewer L N, Kotula P G and Michael J R 2008 *Ultramicroscopy* **108** 567-578
[8] Wilkinson A J, Collins D M, Zayachuk Y, Korla R and Vilalta-Clemente A 2019 *Ultramicroscopy* **196** 88-98
[9] McAuliffe T, Foden A, Bilsland C, Daskalaki-Mountanou D, Dye D and Britton T 2020 *Ultramicroscopy* **211** https://doi.org/10.1016/j.ultramic.2020.112944
[10] Jolliffe I T 2002 *Principal component analysis. 2nd edition*. (New York, NY: Springer)
[11] de la Peña F, Ostasevicius T, Fauske V T, Burdet P, Jokubauskas P, Nord M, Sarahan M, Prestat E, Johnstone D N, Taillon J, *et al.* 2017 *Microsc. Microanal.* **23** 214-215
[12] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al.* 2011 *J. Machine Learning Res.* **12** 2825-2830
[13] Ånes H and Bergh T 2020 *KIKUCHIPY v0.1.2* https://github.com/kikuchipy/kikuchipy
[14] Villars P and Calvert L 1985 *Pearson's handbook of crystallographic data for intermetallic phases*. (Metals Park, OH: American Society of Metals)
[15] Colville A A and Ribbe P H 1968 *Amer. Mineralogist* **53** 25-37
[16] Powell M J D 2009 *The BOBYQA algorithm for bound constrained optimization without derivatives*. Report DAMTP 2009/NA06. (Cambridge, UK: University of Cambridge) 26-46
[17] Hielscher R and Schaeben H 2008 *J. Appl. Crystallography* **41** 1024-1037
[18] Hansen L T, Jackson B E, Fullwood D T, Wright S I, De Graef M, Homer E R and Wagoner R H 2017 *Microsc. Microanal.* **23** 460-471
[19] Jackson M, Pascal E and De Graef M 2019 *Integrating Mater. Manufact. Innovation* **2** 1-21