ARTICLE

Check for updates

Identifying chemically similar multiphase nanoprecipitates in compositionally complex non-equilibrium oxides via machine learning

Keyou S. Mao [1,2✉], Tyler J. Gerczak[1✉], Jason M. Harp [1], Casey S. McKinney[1,3], Timothy G. Lach [1], Omer Karakoc[1], Andrew T. Nelson[1], Kurt A. Terrani[1,4], Chad M. Parish[1✉] & Philip D. Edmondson[1✉]

Characterizing oxide nuclear fuels is difficult due to complex fission products, which result from time-evolving system chemistry and extreme operating environments. Here, we report a machine learning-enhanced approach that accelerates the characterization of spent nuclear fuels and improves the accuracy of identifying nanophase fission products and bubbles. We apply this approach to commercial, high-burnup, irradiated light-water reactor fuels, demonstrating relationships between fission product precipitates and gases. We also gain understanding of the fission versus decay pathways of precipitates across the radius of a fuel pellet. An algorithm is provided for quantifying the chemical segregation of the fission products with respect to the high-burnup structure, which enhances our ability to process large amounts of microscopy data, including approaching the atomistic-scale. This may provide a faster route for achieving physics-based fuel performance modeling.

[1] Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [2] Present address: FAMU-FSU College of Engineering, Florida State University, Tallahassee, FL 32310, USA. [3] Present address: Department of Materials Science and Engineering, University of Florida, Gainesville, FL 32611, USA. [4] Present address: Ultra Safe Nuclear Corporation, Seattle, WA 98199, USA. ✉email: kmao@eng.famu.fsu.edu; gerczaktj@ornl.gov; parishcm@ornl.gov; edmondsonpd@ornl.gov

Understanding the evolution of the nuclear fuel in service is critical to ensuring a reliable environment for geological storage sites[1], improving the fuel efficiency and safety as the operation is pushed to higher burnups[2,3], and providing a realistic low carbon emission energy solution[4–6]. The current fleet of light-water reactors (LWRs) utilizes oxide fuel such as UO_2. The burned nuclear fuel contains multiple phases, including fission-product phases with complex chemistry. These features can be represented by noble gases (e.g., Kr and Xe) and metallic fission products (e.g., Mo, Tc, Ru, Rh, and Pd), whereas other fission products form oxides such as Cs, Ba, Zr, or lanthanides and transuranic elements[1,7–9]. Since the temperature decreases from the pellet center to the periphery, the generated fission products distribute heterogeneously in the UO_2 matrix. The fission gases migrate into the defect structures because of their low solubility and tend to coalesce in the grain interior as intragranular bubbles or at the grain boundaries and triple junctions as intergranular bubbles. When the fission gas moves into the fuel cladding gap, the released fission gas decreases the heat transfer in the gap, leading to an increased central fuel pellet temperature with time. Such a process accelerates the swelling, cracking, and mechanical degradation of the fuel and plenum. However, it is demanding to interpret the structure and composition of complex fission products, especially those with elements that have overlapping X-ray lines, which is common because of the high fraction of fission-product yield with similar chemistries. For instance, to accurately predict the thermal conductivity of a UO_2 fuels along its lifetime, the mechanistic fuel-performance modeling is essential to account for the impact of gaseous fission products, where nanoscale point defects like Xe atoms influences phonon scattering, as well as the restructured grains, affect the thermal transport of the high-burnup microstructure[10,11]. The current fuel-performance modeling software including the MARMOT[12] and BISON[13] code incorporated in the MOOSE Simulation Environment[14] are seeking to improve the mesoscale simulation module to a full thermal conductivity model with coupled radiation damage such as swelling and fission gas release. Thus, understanding these complex fission-product release behavior processes requires transformational science through detailed experiments and simulations at multiscale, coupled with physical/chemical modeling[15]. This approach will eventually provide such knowledge to the engineering-scale domain.

The multiscale modeling capabilities must reflect all physical effects in the experiments. The general challenge of characterizing fission products in irradiated nuclear fuel—in which large fractions of fission products are present—lies in gathering of accurately combined comprehensive crystallography data, elemental distribution, and response to the service environment captured from experiments. Modern characterization of irradiated nuclear fuels adopts a multifaceted approach. A combination of advanced experimental techniques is employed, including two-dimensional (2D) electron microscopy such as scanning electron microscopy (SEM) with focused ion beams (FIBs), transmission electron microscopy (TEM), scanning TEM (STEM), electron energy loss spectroscopy (EELS), energy-dispersive spectroscopy (EDS), and X-ray diffraction (XRD) in various forms of fuel systems. These fuel systems include metallic fuels[16–18], MOX fuels[19–22], uranium oxide fuels[23–25], and tristructural isotropic fuels[26,27]. The advent of three-dimensional (3D) characterization reconstruction in materials science[28] facilitated advancements in fuel characterization techniques such as atom probe tomography (APT)[29,30] and synchrotron X-ray tomography (XRT)[31]. Although some 3D reconstruction provides new insights into the local microstructure, the inability to cover a sufficient volume of the fuel section of interest, high-cost setups with destructive methods (FIB slicing or APT), and labor-intensive analysis process make it less

representative of the macroscopic level; other 3D imaging techniques such as XRT can cover large volumes at the expense of resolution, while revealing smaller-length scale details in the structure commonly requires longer data acquisition time[32].

With respect to the X-ray mapping inside a SEM or STEM using EDS systems, the elemental distributions of materials can be probed at high count rates by the large-area high-throughput silicon drift detector technologies, resulting in full-spectrum X-ray imaging (XSI) at each pixel instead of the conventionally recorded X-ray intensity maps versus position[33]. XSI maps can be analyzed through conventional curve-fitting approach based on a sufficient signal-to-noise ratio in each pixel, where the background is subtracted for deconvolution of the overlapping signals[34,35]. However, the volume of the data produced by modern-day microscopy can be up to 300 GB per hour[36]. In order to effectively and efficiently interpret the data unbiasedly, augmented data analytics algorithms are required to be incorporated into the workflow by a human analyst, which is often found out that rich details invisible to manual analysis can be revealed by these advanced methods[37–39]. Recent advances in ptychography methods such as four-dimensional (4D) STEM and precession electron diffraction (PED) have generated big data, and these advances are critical for the development of efficient analysis tools to identify precise structures and extract useful features from the big data[36,40]. In the artificial intelligence (AI) and computer vision community, machine learning (ML) has been applied to many problems: classification, image recognition, segmentation, tracking, and reconstruction[41]. Their research and development on advanced ML algorithms promoted open-source ML packages such as Scikit-learn[42], and deep learning (DL) packages like TensorFlow and PyTorch[43,44]. There are three primary barriers to implementing these techniques into the microscopy and materials science fields: (1) the cost of material preparation and infrastructure for data collection is high, limiting the available datasets for post analysis; (2) prior knowledge in materials science and relevant disciplines is needed to accurately label the data; and (3) there is no standardized platform that can be used instead of curve-fitting models to translate raw data of various forms into organized workflows. A framework for crystallographic determination through DL on massive diffraction datasets was developed recently by Aguiar et al.[45]. Their convolutional neural networks enable the augmented analysis of diffraction patterns without stored data but have not yet addressed unsupervised ML to either pre-process the images for training or post-process the images through clustering methods. With respect to nuclear fuel analysis, the structure of fission products is usually unknown; thus, the indexing of the diffraction pattern is unlike that of the supervised ML method on known objects, in which the potentially matched patterns can usually be found in a crystallography structure library such as Pearson's Crystal Data[46], the Crystallography Open Database[47], and the Materials Project[48]. The scale of fission products can expand from nanoscale to macroscale, and for the nanophase features, selected area electron diffraction (SAED) patterns can be challenging to acquire. For applicability at the engineering scale, modeling and validation of the computer codes require the microscopy data to cover as large an area as possible on the fuel section. STEM-EDS can balance these requirements and provide sufficient data for the irradiated fuel characterization, but the identification of the X-ray peaks can be difficult when the fission products are too close to each other. Identification can also be difficult because of noise-caused artifacts resulting from sample preparation, which is a significant challenge for spent nuclear fuel where remote handling can make even standard metallographic polishing methods difficult. Moreover, STEM-EDS of these materials is particularly challenging because of the large number of elements present caused by the burnup and
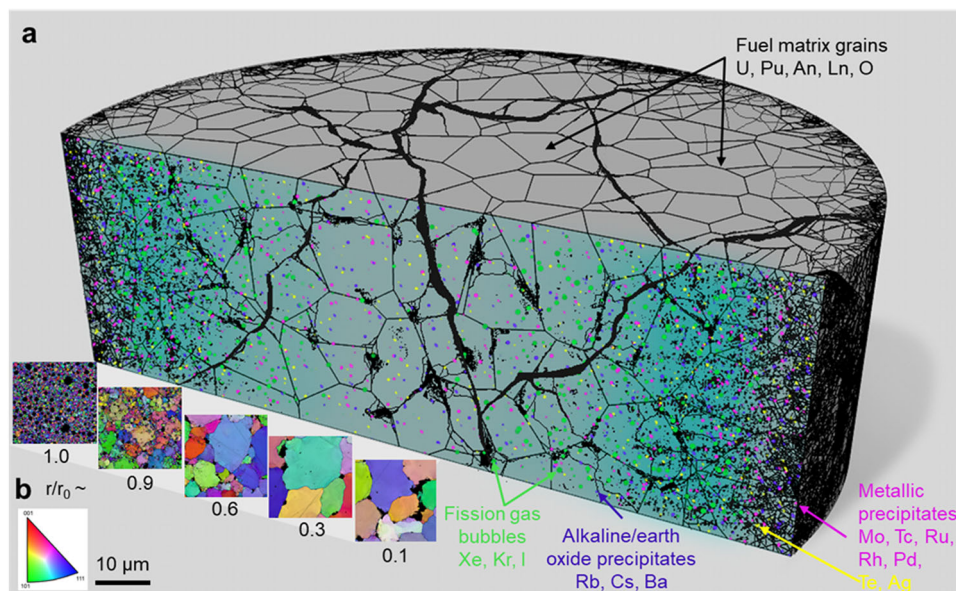
Fig. 1 Schematic cross-sectional view of the spent nuclear fuel microstructure. a Magenta and yellow elements are metallic precipitates formed by fission products, whereas green elements are bubbles or pores formed through fission gas release such as Xe, Kr, I. The fission products can segregate in the crack and decorate both the surfaces of the grain boundaries and grain interiors, resulting in further intragranular/intergranular bubble swelling. The grain restructuring (subdivision) occurs at the transition region and fully develops into submicron grains toward the fuel rim. **b** Selected EBSD scans in a fuel from the Limerick NPP are provided as an example to demonstrate the grain structure at five different radial positions. The transition region has a combination of both large and submicron grains, and the rim indicates the fully developed HBS.

decay of the material during service. Given that a large number of elements present are closely located on the periodic table, the chemical characteristics may be very similar. By extension, the characteristic X-rays that make elemental discrimination possible may be close to the energy resolution of the detectors for many of the elements present. For example, one possible decay chain is the $Ba \rightarrow Cs \rightarrow Xe$ transition that can be used for forensic examination of the fuels, but these elements are also difficult to discriminate because of their similar characteristic X-ray energies.

To accelerate this characterization cycle, we developed an ML-driven approach to enhance the data mining of contemporary microscopy data on irradiated fuels. Specifically, this method was demonstrated on high-burnup fuels and their fission products, and it also revealed the metallic precipitates and fission gas at the nanoscale. The highly irradiated commercial spent nuclear fuel samples were examined in different regions through state-of-the-art microscopy. The subsequent high-quality visualization through combined advanced ML algorithms captured the unique nanoclusters at different sinks, such as the metallic precipitates and grain boundaries. The principal component and clustering analyses show the detailed chemistry, phase compositions, and tendency of segregation in a large dataset composed of random precipitates. With this advanced approach, the current nuclear fuel characterization cycle could be significantly improved leading to accelerated innovative fuel design based upon reliable experimental data, with great potential to reveal more nanoscale fission-product release behaviors, contributing to a carbon-neutral option with "ultra-high-burnup" fuels.

Results

Principal component analysis of restructured fuel. A schematic representation of the spent fuel structure is provided in Fig. 1a[49–52]. Fission gas bubbles (green) and metallic fission-product precipitates (magenta/yellow) influence the fuel performance. The fission products can segregate in the grain boundaries and decorate both the surfaces of the grain boundaries and grain interiors,

resulting in further intragranular/intergranular bubble swelling, and they develop into larger pores relative to the submicron- or micron-sized pores/bubbles. Cracks form both intergranular and transgranular. The larger grains restructure into submicron grains toward the pellet rim as the high-burnup structure (HBS) forms. Figure 1b shows an example of the grain structure transition in spent fuel from the Limerick Generating Station's boiling water reactor (BWR) nuclear power plant (NPP) with an average burnup of 56 MWd kgU^{-1}. Selected cross-sectional electron backscatter diffraction (EBSD) maps at five different radial positions demonstrate the grain structure as a function of radius. In the grain structure, the central region has large (>10 μm) grains, the transition region has a combination of both large and submicron grains, and the rim indicates the fully developed HBS. The grain structure resembles fuel from the H.B. Robinson pressurized water reactor (PWR) NPP[23] and has a similar burnup level.

The high-burnup UO_2 microstructural evolution can be characterized based on the variable chemistry of fission-product and fission gas features along the radius of the cross-section. An area was selected in the Limerick NPP spent fuel transition region ($r/r_0 \sim 0.9$, where r_0 is the radius of the pellet and r is the position along the radius from which the data were acquired). This region combined both unrestructured and restructured grains, shown by the bright-field STEM (BFSTEM) in Fig. 2a. Bright-field TEM (BFTEM) using the Fresnel contrast method was applied to image the nanoscale Xe bubbles in the unrestructured region enclosed by the red box in Fig. 2b. Then, a transmission Kikuchi diffraction (TKD) map was used (Fig. 2c) to confirm the combination of both unrestructured grain (purple) and its surrounding submicron restructured grains: only nanoscale Xe bubbles exist in the unrestructured grain. Finally, Fig. 2d displays SAED patterns, which shows the matrix UO_2 on zone axis 112. The minor diffraction spots indicate the small metallic fission-product precipitates (<70 nm). Examples were given to fit possible phases: body-centered cubic (BCC) β-Mo (Tc, Ru) on zone axis 110, face-centered cubic (FCC) α-Pd (Ru, Rh) on zone axis 112, and
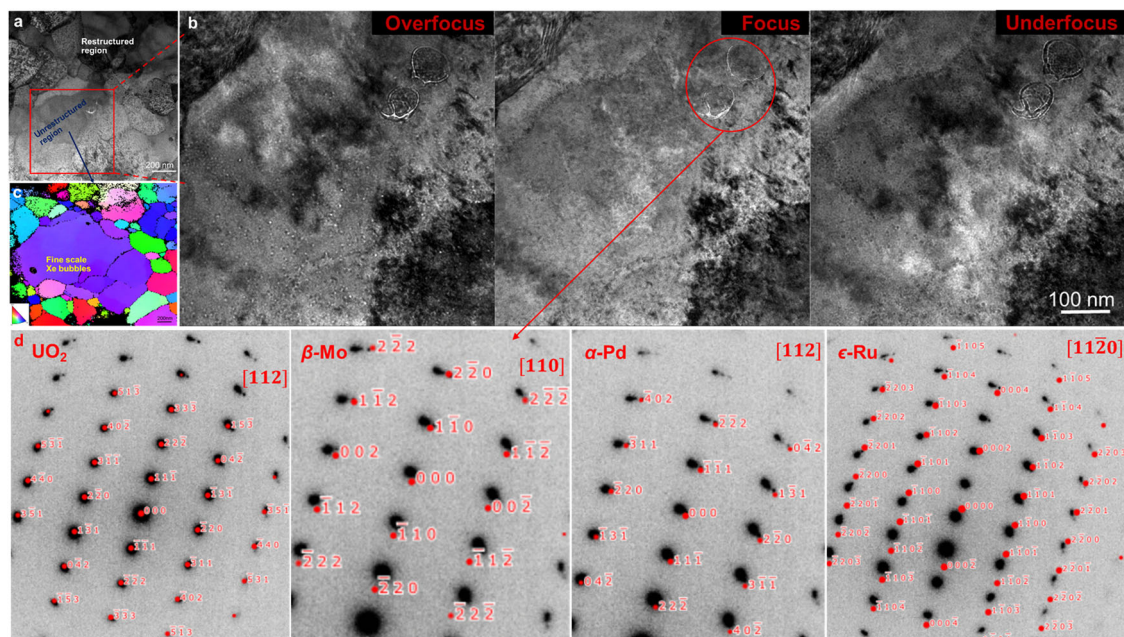
Fig. 2 Conventional approach to determine the crystal structure of fission-product precipitates and Xe bubbles. a BFSTEM shows the microstructure of the transition region of the Limerick fuel. **b** Conventional TEM using the Fresnel contrast method can image the Xe bubbles but requires a longer time of operation. **c** A TKD map confirms the combination of both unrestructured (purple) grains and its surrounding submicron restructured grains, whereas only nanoscale Xe bubbles exist in the unrestructured grain. **d** SAED patterns show the matrix UO_2 on zone axis 112. The minor diffraction spots are difficult to index because of the unknown structure of the small metallic fission-product precipitates (<70 nm). Examples were given to fit possible phases from literature[25,53,54]: BCC β-Mo (Tc, Ru) on zone axis 110, FCC α-Pd (Ru, Rh) on zone axis 112, and HCP ϵ-Ru (Mo, Tc, Rh, Pd) on zone axis 11$\bar{2}$0.
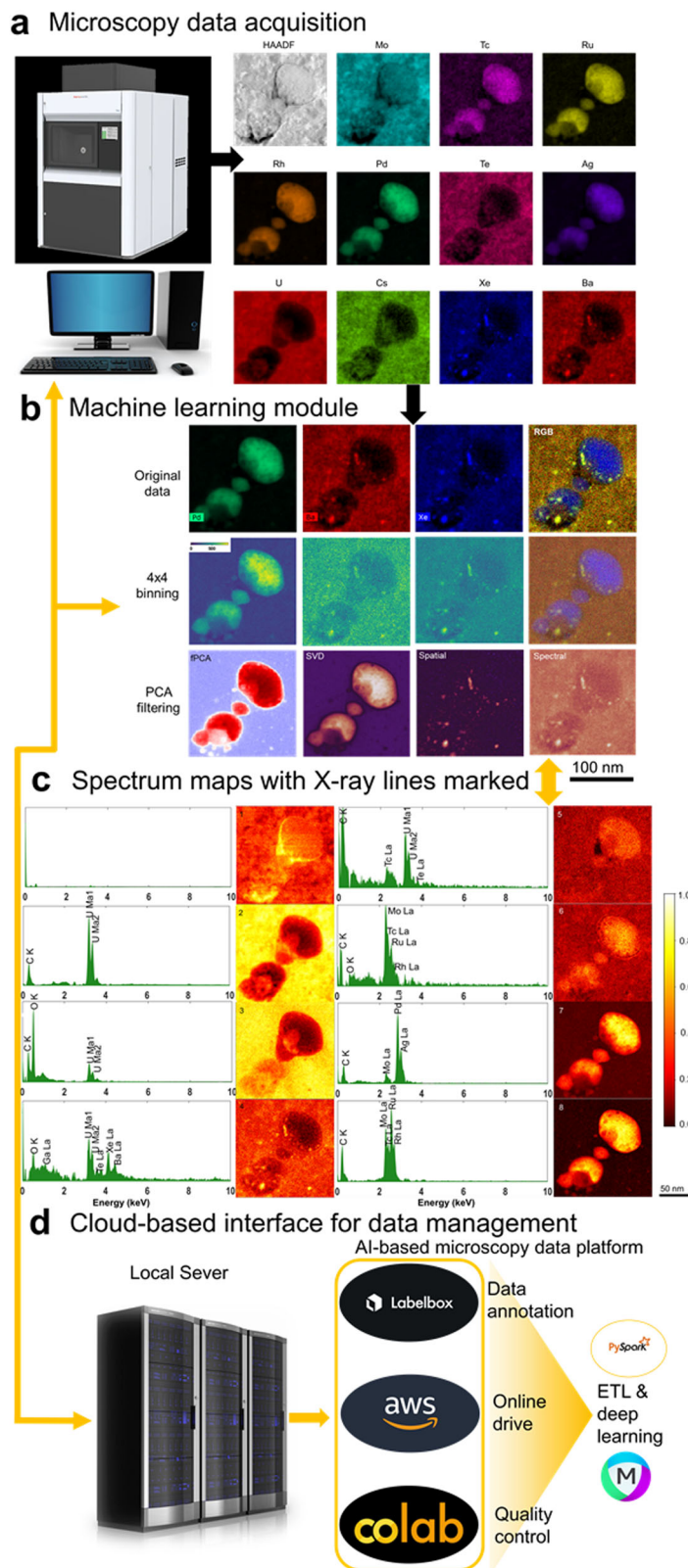
hexagonal close-packed (HCP) ϵ-Ru (Mo, Tc, Rh, Pd) on zone axis 11$\bar{2}$0. However, it is challenging to accurately index these precipitates because of the following reasons: (1) the composition and crystallography of these fission products[25,53,54] are intricate in nature; (2) there is limited capability to substantially thin a specific region of the highly radioactive, porous fuel specimen for high-quality high-resolution TEM (HRTEM) and fast Fourier transform (FFT); and (3) even though assuming a good-quality thin specimen is achievable, it is inefficient to locate, record, and determine the structure of multiple precipitates through time-consuming tilting, correction, and indexing. Therefore, we propose an advanced approach enhanced by ML to help obtain useful information, especially about the chemistry of these small fission-product precipitates, nanoscale Xe bubbles, and potential nanoclusters that are unable or difficult to be resolved in a conventional TEM.

First, we focused on the metallic fission-product precipitates shown in Fig. 2 using targeted STEM-EDS at nanoscale resolution. A demonstration of the ML-driven workflow is shown in Fig. 3. This approach starts with the input of a high-quality raw STEM-EDS image (Fig. 3a). The raw data, for example, cannot distinguish the overlapping of Ba and Xe and artifacts. The data were analyzed using the optimally scaled principal component analysis (PCA) approach described by Keenan et al.[55] Raw data are scaled for Poisson noise and subjected to singular-value decomposition (SVD), and the model is truncated to only relevant components and the translated data then returned from Poisson space to its original space. This results in a loss of orthogonality, so a reorthogonalization is applied. From this PCA-like model, matrix rotations (such as Varimax) can be applied to convert either the abundance maps or spectral endmembers into a simple representation (Fig. 3b). More details about this process can be found in the scree plot (Supplementary Fig. 1), SVD before truncation channels (Supplementary Figs. 2

and 3), functional PCA (FPCA) (Supplementary Figs. 2 and 3), and Varimax rotations (Supplementary Figs. 6–9).

A demonstration of the unsupervised ML approach for the labeled spectrum data is given in Fig. 3c. The hyperspectral data were converted through an eight-component decomposition to reveal the detailed clustering. Each component represents a calculated score image applied by the PCA. Components 1–3 show the U element loading spectra with a strong score image away from the precipitates; this is the fuel matrix. Component 3 shows the same U elemental lines, but it has strong O-K lines and weak U-M lines compared to component 2. The combination of 2 and 3 demonstrates the self-absorption of the soft (lower energy) X-ray M lines together, thus demonstrating the U-rich matrix. Component 4 shows Xe-L and Ba-L lines together, which is a strong indication of Xe and Ba nanoclustering near the precipitates. Component 5 shows high C and U signals but a weak C signal in the lower-left precipitate. Component 6 indicates a strong Mo signal, along with Tc, Ru, Rh in all four fission-product precipitates close to the β-Mo (Tc, Ru). Component 7 shows an even stronger Pd contribution, as well as Ag, C, and Mo in the center of the upper-right precipitate accompanied by small clusters near the α-Pd (Ru, Rh). Component 8 provides stronger Ru content with weaker Mo, Tc, and Rh enrichment on the two larger precipitates, and it provides nanoclusters distributed in the matrix, indicating similarity in ϵ-Ru (Mo, Tc, Rh, Pd).

We consider that the scale of the data generated by the modern-day microscopes and provide a strategy for cloud-based data management[56,57]. The collected data in the local server can be transferred to online storage through an Amazon Web Service (AWS) cloud server. The microscopic images can then be annotated through Labelbox, which is used to train DL to obtain valuable statistics from large datasets. Extract, transform, load (ETL) is also achievable via the PySparks platform to gather data

a Microscopy data acquisition

b Machine learning module

c Spectrum maps with X-ray lines marked

d Cloud-based interface for data management

from multiple resources for in-depth data-driven analysis. Google Colaboratory can serve as a quality control tool for tracking the image analysis such as further segmentation and reconstruction as each user could implement their own scripts with user-defined AI algorithms. This dynamic structure can improve the quality of data analysis and can be implemented into the data acquisition process. Users can specify their own scripts to manage their workflows using such an AI-based microscopy data platform. Such a platform greatly expands the compatibility of microscopy data heterogeneity from different users and can provide a flexible channel to compare various forms of data on a consistent path. This also allows users to incorporate different codes/algorithms from GitHub repositories for either DL or other types of AI-based analyses, for instance, image analysis software MIPAR[58]. Another

Fig. 3 Unsupervised ML approach reveals detailed decay product nanoclusters in the vicinity of the metallic fission-product precipitates. a Data acquisition using the state-of-the-art Talos F200S G2 TEM. Also shown are the raw STEM-EDS intensity maps of precipitates in Fig. 2, in which overlapping of Ba and Xe could not be differentiated and were sent into the data pipeline processing modules for post analysis. **b** The data were processed through a multi-layered ML module structure. The method demonstrated here used spectrums (X-ray energy, keV) with an eight-component decomposition. PCA deconvolution of the spatial and spectrum distributions rotated to the optimally scaled simplicity. An example of noise-filtered maps after a 4 × 4 data binning and PCA includes fPCA, SVD, and spatial/spectral simplicity. An RGB color mix of Ba, Pd, and Xe[60] is shown. X-ray spectrum peak identification can be achieved through the nonnegative matrix factorization (NMF) method[61,62]. **c** Spectrum maps with X-ray lines marked. The y axis displays the intensity of the signal (arbitrary units/counts), and it should be noted that the same magnitude is not displayed in each sub-figure and the maximum intensity in each plot is different. The selected area is in the transition region (r/r_0~0.9) of the Limerick BWR fuel rod of average burnup at ~56 MWd kgU^{-1} (~5-year service in an NPP reactor). The results indicate the association of the Xe and Ba clusters with the metallic precipitates (Mo, Tc, Ru, Rh, Pd, and Ag). Since no mobile Cs were detected, the coexistence of the Xe and Ba clusters indicates that the Ba is a decay product arising from the decay chain ^{138}Xe (14-min half-life)→^{138}Cs (33-min half-life)→^{138}Ba from neutron inventory analysis. **d** Overview of cloud-based data management strategy. "Amazon Web Services, the "Powered by AWS" logo, are trademarks of Amazon.com, Inc. or its affiliates in the United States and/or other countries.".

recent advance of automated mapping software enables large-region acquisition of montages of XSIs interrogated by computer-controlled stage motion, which provides both big datasets while maintaining fine pixel pitch[59]. Thus, the very large datasets can span multiple orders of magnitude of spatial scale, making it difficult for the human analyst to interrogate comprehensively. Within a constant total time, because the larger-areal mapping reduces the time per tile, degradation of signal-to-noise ratio may hide low-intensity features under the noise floor. Here, our ML approach in Fig. 3 provides an opportunity to leverage highly optimized algorithms both computationally and memory-efficiently on the sparse data processing based on the intrinsic sparsity of XSIs.

Improved visibility of Xe bubbles through machine learning. The migration of Xe gas is affected by fission gas sinks[63–65], so targeting specific locations in the experiment is important because of the expectations of the nanophase fission gas behaviors. However, the nanoscale Xe bubbles in the restructured region shown in Fig. 2 are difficult to locate in a large area with a conventional TEM approach. The quantitative statistical analysis on these nanoscale Xe bubbles is limited by high-angle annular dark-field (HAADF) comparison across multiple regions and could miss information on even finer bubbles. Here, we reapplied the ML-driven approach demonstrated in Fig. 3 to reveal the intragranular nanoscale Xe bubbles shown in Fig. 2b. PCA results show both precipitates and Xe bubbles. Detailed X-ray intensity mapping of the Xe (Fig. 4) improves the visibility of fine Xe bubbles, which are invisible when using BFTEM in Fig. 2a. Cs and Ba are not prevalent in the matrix. Figure 4b shows a higher magnification deconvolution map (a spectral simplicity–rotated abundance map and a spectral endmember pair) of intragranular Xe bubbles in another restructured grain of the Limerick fuel. In this particular instance, the PCA has been applied to greatly improve the relatively noisy and sparse dataset, into a series of low-rank, low-noise, image, and spectra vectors. Reducing the data into these components vastly improves the visibility of key features. In this instance, the PCA analysis revealed three primary components representing the three strongest spectrum-image pairs within this particular dataset. The three-component analysis revealed overlapping of Xe bubbles with the fission products. Each component was marked as endmember #0–#2, and an abundance map accompanies each endmember. Each X-ray (keV) peak labeled via the multivariate curve resolution-alternating least-squares method[66] allows efficient identification of nanoscale Xe bubbles, and it also displays finer details of fission-product nanoclusters (Fig. 4c). The C–K and O-K lines are weak across the three components. Component 1 shows the U-M lines, which are weaker at the Xe bubbles than the matrix. Component 2 shows the stronger Mo-L and Ru-L signal and weaker Tc, Rh, Pd, and Ag signals, implying the combination of the β-Mo (Tc,

Ru) and ε-Ru (Mo, Tc, Rh, Pd) phases. Component 3 shows stronger Xe-L lines than the Mo–Tc–Ru–Rh–Pd–Ag peaks. This demonstrates the fission-product nanocluster segregation is associated with nanoscale Xe bubbles. We also examined the HBS; no intragranular Xe bubbles were found. This ML-aided algorithm improves the accuracy and speed of locating fine Xe bubbles smaller than 5 nm.

To validate our ML-aided EDS approach, we extended the algorithm to the HBS in the H.B. Robinson fuel with a higher average burnup at approximately 72 MWd kgU^{-1}. Like in the HBS of the Limerick fuel, no intragranular Xe bubbles were observed. However, nanoscale Xe bubbles still exist intergranularly at the grain boundary in the HBS. The advantages gained from using the ML algorithms were validated (Fig. 5), and conventional TEM using the Fresnel contrast method was unable to capture finer features at a random high-angle grain boundary. After the ML reconstruction of the STEM-EDS data, intragranular Xe bubbles overlapping with the fission-product clusters were observed (Fig. 5b). The nanoscale Xe bubbles (≤4 nm) are presented as color mix maps in Fig. 5c, and the spectrum is labeled at low- (0–10 keV) and high-energy (10–20 keV) ranges. The ML decomposition of the X-ray spectrum at the low- and high-energy ranges in Fig. 5d demystifies the fission-product nanoclusters and Xe bubbles trapping at the grain boundary. In the low-energy range below 5 keV, the fission product precipitates are rich in Mo-L peak, and the Tc-L and Ru-L peaks are lower. The Xe bubbles are reflected by the Xe-Lα, accompanied by Xe-Lβ$_1$ overlapped with U-M$_2$N$_4$ and minor Xe-Lβ$_2$. The matrix is dominated by the U-Mγ line. At the high-energy range above 18 keV, a strong Mo-Kβ signal appears, along with Ru-Kα, Tc-Kα, Pd-K, and U-Lγ$_1$ peaks. The spectrum analysis showed that the fission product is the β-Mo (Tc, Ru) phase. In this scenario, the Xe-L line readings are lower than the spectra loadings of fission product precipitate L lines. However, the nanoscale Xe bubbles can be recognized with improved visibility over the conventional BFTEM and STEM methods.

Segregation of fission-product precipitates through data mining. In addition to the qualification of fission products, ML can be applied to quantify the composition of each individual fission-product precipitate. Figure 6a displays the composition of small-scale metallic fission-product precipitates (<100 nm) in the HBS as a heatmap along the radial position of the Limerick fuel segment. The heatmap was calculated from normalized atomic percent (at. %) for Xe, U, Te, Tc, Ag, Ru, Rh, Pu, Pd, Mo, Cs, and Ba with acquired data from STEM-EDS maps. Red indicates high levels of concentration, whereas blue and green indicate low concentration levels. The precipitates primarily consist of Mo, Tc, Ru, Rh, and Pd, with minor Ag contributions; Te, Ba, Cs, and Xe are present in lower concentrations. The phase identification of
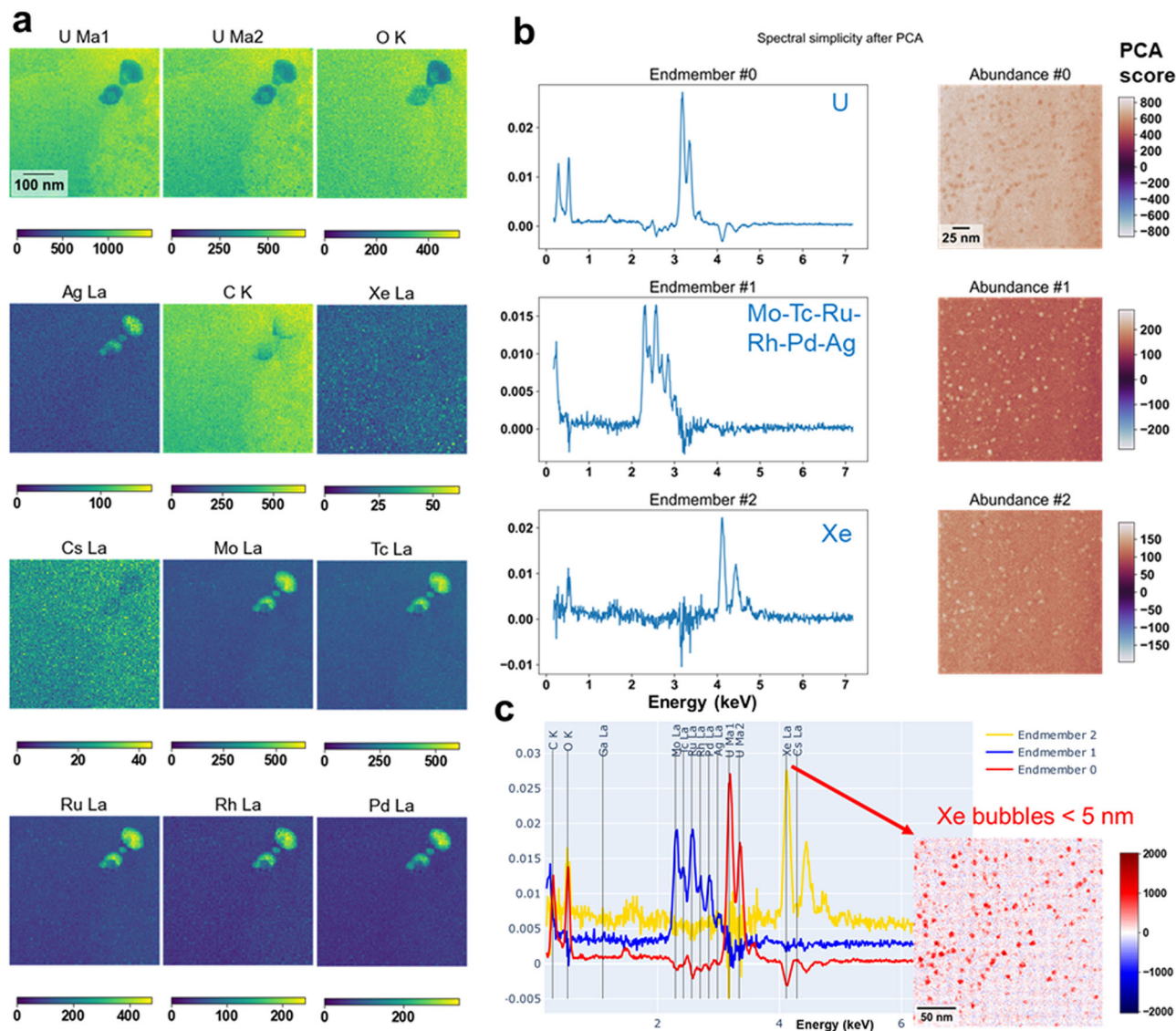
Fig. 4 Machine-learning–aided microscopy improves the accuracy and speed of locating intragranular nanoscale Xe bubbles. a STEM-EDS intensity maps after 4 × 4 data binning show the intragranular Xe bubbles in the transitional region. These STEM-EDS maps intensity maps are taken from the same region presented in Fig. 2a. **b** Spectral simplicity rotation after PCA representation (X-ray energy in keV) shows the overlapping of Xe bubbles with the metallic fission products in the restructured grain. **c** Spectrum simplicity after PCA allows efficient identification of nanoscale Xe bubbles; finer details of nanoclusters are also displayed. Application of ML to the STEM-EDS spectrum images improves the visibility of fine Xe bubbles, which cannot be seen in Fig. 2a.

these fission-product precipitates was not apparent, so clustering analysis was applied on the chemical mapping based on the *k*-means clustering algorithm. One of the advantages of clustering over bilinear factor models is that clustering can have more clusters than the chemical rank (i.e., "parsimony restriction," Stork and Keenan)[67]. Regarding the data mining and informatics in Fig. 6, the clustering algorithms are performed on the spectral-only data from the standardless Cliff–Lorimer method[68]. PCA on the elemental concentration determines the sizes of five clusters (blue: cluster 1; red: cluster 2; orange: cluster 3; green: cluster 4; purple: cluster 5) for the k-means clustering method. The two-component (Fig. 6b) and three-component (Fig. 6c) PCA decomposition were plotted by using the *k*-means clustering method.

The relationship between each fission-product element and the U content (at. %) was then plotted, shown in Fig. 6d. The major elements Mo, Tc, Ru, Rh, and Pd indicate a strongly negative correlation to the U content. Cluster 1 in blue is present in all the

plots with U contents lower than 20 at. %; Mo, Tc and Pd are major components, and Tc and Rh are minor, indicating a combination of three phases: β-Mo (Tc, Ru), ε-Ru (Mo, Tc, Rh, Pd), and α-Pd (Ru, Rh). Cluster 2 in red represents U content ~20–60 at. %; the major contents are Ru and Mo with minor Pd, Tc, and Rh. The significant change in the Pd concentration indicates the separation of the ε and α phases. Cluster 3 in orange shows Pd depletion, indicating a decrease in the α phase; part of cluster 3 (>1 at. %) is related to the higher Xe concentration. Cluster 4 in green and cluster 5 in purple are majorly influenced by Mo. The decrease in Ru represents ε phase depletion, and the increased concentration of Te, Cs, and Ba in cluster 4 indicate a possible decay pathway of the Te fission product[25]. Ag is only prevalent in clusters 1 and 2 at the radial position ≥0.9.

Figure 6e provides the scatter plot of the ternary phase diagram for each predicted fission-product phase, and individual metallic precipitate compositions were derived from clustering analysis of the STEM-EDS dataset. The α phase diagram distinguishes cluster
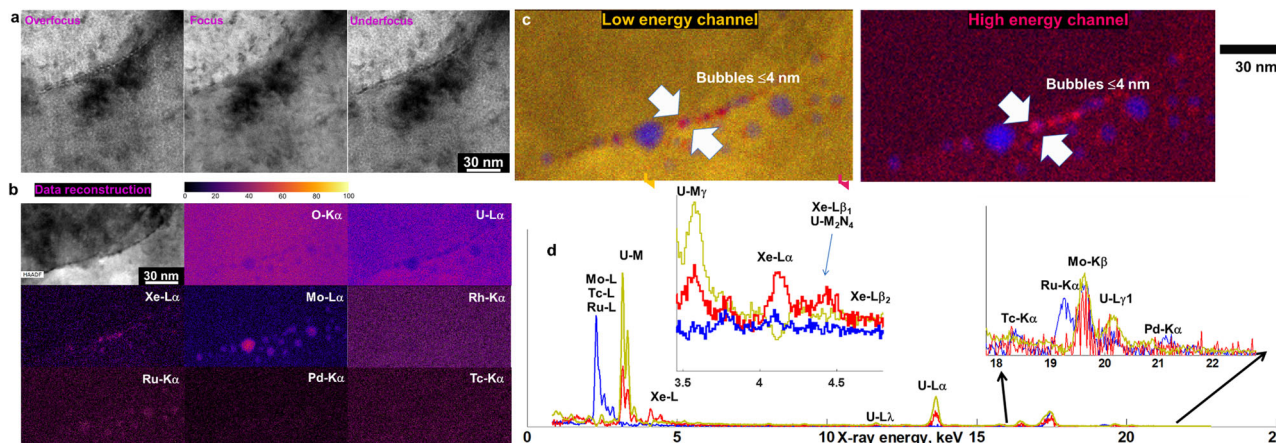
Fig. 5 Unsupervised ML improves the visibility of nanoscale Xe bubbles at the grain boundary. a Conventional TEM using Fresnel contrast method performs poorly in terms of showing fine features at the grain boundary. The data were acquired from the HBS of the H.B. Robinson PWR fuel rod with average burnup at ~72 MWd kgU^{-1}. **b** ML-aided reconstruction of the STEM-EDS data shows the presence of intergranular Xe bubbles overlapping with the fission-product clusters. **c** Combined low-/high-energy channel color mix maps reveal the fine-scale Xe bubbles (≤4 nm). **d** ML decomposition of the X-ray spectrum at the low- and high-energy ranges demystifies the fission-product nanoclusters and Xe bubbles trapping at the grain boundary. Raw STEM-EDS map can be referred as Supplementary Fig. S10.

1 from the others, showing that it has a higher average Pd percentage. The divergence of the scatter data indicates that the α phase may not be a predominant solid solution path. The β phase diagram indicates a strong concentration of the data, and each grouped cluster is closer in the composition spacing, which indicates that most of the fission products are β phase. The ϵ phase diagram resolved the uncertainty in the α phase diagram, indicating that cluster 1 is closer to the ϵ phase, with a Mo content less than 50 at. %. Cluster 2 partially shows the nature of ϵ phase, with a (Ru + Tc) content higher than 75 at. %. Clusters 4 and 5 may also belong to the ϵ phase, with Mo contents lower than 50 at.%. In all the three phase diagrams, cluster 3 is less prevalent than the other clusters, which reverifies the relationship between the clusters and Xe established by the regression plot in Fig. 6d.

Discussion

The workflow in Fig. 3 can provide higher spatial/spectral precision from current microscopy datasets. This method reduces the high-dimensionality EDS data into spectral endmembers and abundance maps. Although the Xe bubbles can be routinely imaged via lengthy combined analysis of TEM and TKD data to correlate nanoscale pores with restructuring (Fig. 2), our ML approach has several benefits with respect to reducing the dimensionality of the EDS data. Our approach increases the characterization speed through the application of ML approaches to provide a robust analysis of sparse and/or noisy data, and to identify spatial-chemical relationships that would not normally be possible under human analysis alone, and moreover, the agglomeration of nanophases such as the formation of precipitates/clusters displays an abundance-represented picture of the localized structure (Fig. 3). Direct evidence of the resolution of Xe bubbles is given in Figs. 3 and 4: the ML separates different phases simultaneously and presents these Xe nanoclusters near the vicinity of the metallic precipitates. We have also extended the algorithms to probe the intergranular bubbles in the HBS of the H.B. Robinson NPP fuel. Other PCA methods can be seen in Supplementary Figs. 5, 7, and 9.

Overall, correlations between the elements in the four larger metallic precipitates and several nanoclusters including Xe and Ba are described in the ML analysis, as depicted in Fig. 3c. Compared with the time-consuming SAED, HRTEM, and FFT methods[19,25–27],

our analysis provides a compact qualitative description regarding multiple precipitates' solid solutions and extends to the nanoclusters unable to be resolved with conventional BFTEM. Further denoising and reconstruction can be obtained through optimally scaled PCA, as well as spatial or spectral matrix rotation to achieve a low-rank description of the EDS data. Figure 3b provides the noise-filtered maps after the PCA and NMF peak finding algorithm was performed. The RGB color mix of Ba, Pd, and Xe highlights the overlapping of Ba and Xe (Fig. 3b). The high-quality, low-rank, noise-filtered results indicate the association of the Xe and Ba clusters with the metallic precipitates (Mo, Tc, Ru, Rh, Pd, Ag).

The challenge of STEM-EDS is to identify chemically distinct elements that have characteristic X-rays that are very similar in energy. For instance, the Lα lines of Xe (4.110 keV), Cs (4.285 keV), and Ba (4.466 keV) are very close to each other. The PCA filtering maps allow masked elements to be identified with greater confidence. Moreover, the chemical reactions that occur during the fission release process can be clarified and separated from the radioactive process. We provided an approach to high-throughput spectrum-image mapping to differentiate the closely neighboring elements in X-ray lines (Fig. 3). The clustering of Xe and Ba, which are chemically dissimilar, is likely the result of the radioactive decay of Xe into stable Ba. The most likely source of Xe decaying into stable Ba is ^{138}Xe (14 min half-life)→^{138}Cs (33 min half-life)→^{138}Ba (stable). It is also possible that Ba could come from the atomic number 137 decay chain, but no Cs were detected in these clusters, indicating that longer-lived ^{137}Cs and other Cs isotopes migrated to other areas of the fuel.

From a science perspective, several interesting observations can be extracted from our ML-enhanced XSI mapping. The following insights could be useful for future guidance on fission gas behavior, HBS formation mechanisms, and nuclear fuel characterization. (1) Unlike the rapid diffusion of larger bubbles/pores in the central fuel region, the diffusion mechanism in the HBS was interrupted. (2) The trapping of the Xe bubbles in the unrestructured grains is distinct from the absence of fine-scale intragranular Xe bubbles in the HBS, which is consistent with the previous observation of the Xe depletion in the HBS[69]. (3) The Xe depletion is related to the accelerated fission gas transport, and the intergranular bubbles grow through a different pathway as a result of the newly formed subgrains acting as high-density heterogeneous nucleation sites. (4) The Xe bubbles located in
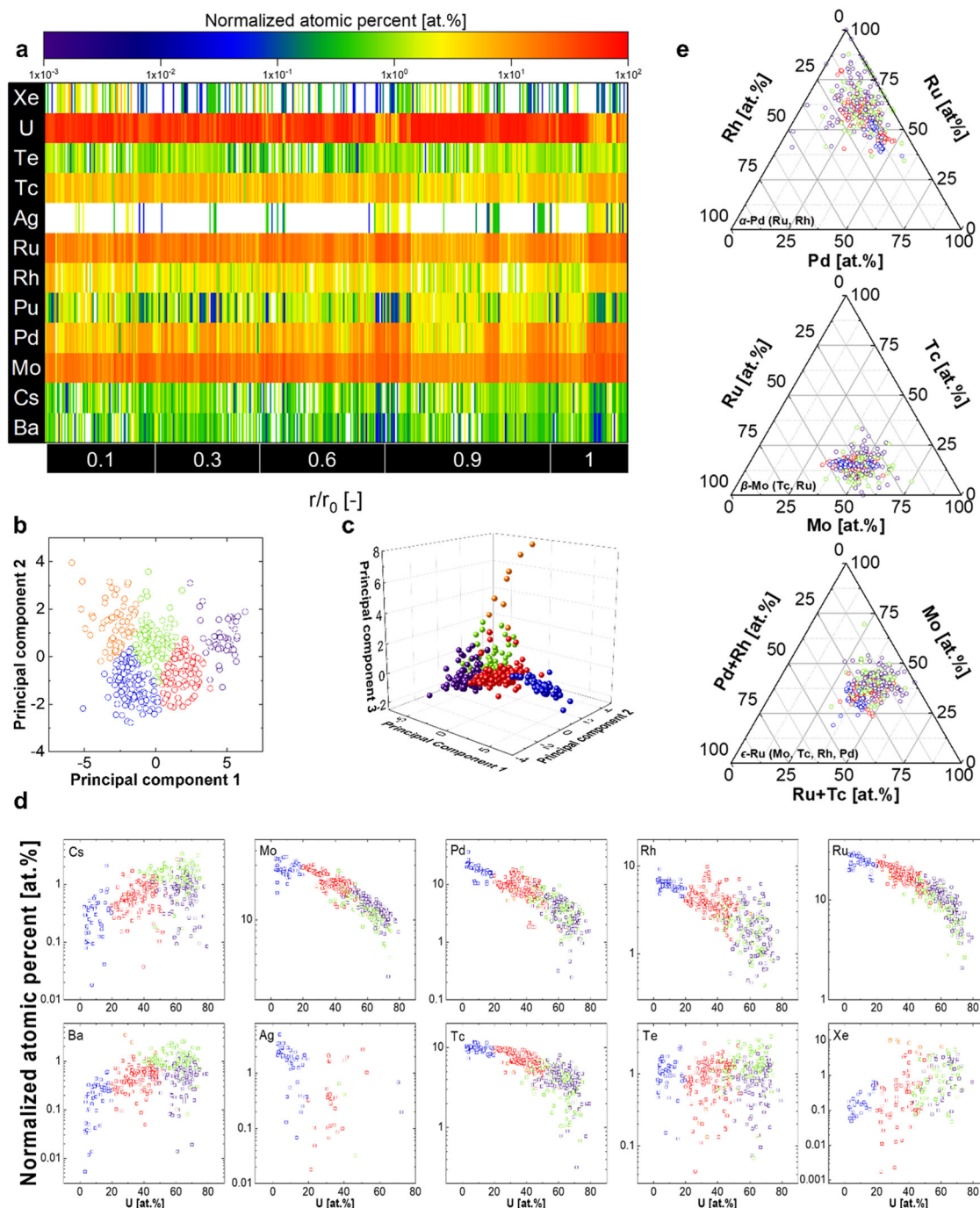
Fig. 6 Materials informatics–driven chemistry analysis on fission-product metallic precipitates along the radial position. a Heatmap calculated from normalized at. % acquired from STEM-EDS maps and radial positions of the metallic precipitates in the Limerick fuel. Red indicates high levels of concentration, whereas blue and green indicate low levels of chemical concentration. The precipitates primarily consist of Mo, Tc, Ru, Rh, and Pd, with minor contributions from Ag. Te, Ba, Cs, and Xe are present as minor components. **b** k-means clustering based on two-component PCA decomposition. **c** k-means clustering based on three-component PCA decomposition. **d** Regression analysis after PCA. A strong correlation was found between the major fission products (Mo, Tc, Ru, Rh, and Pd) and U. **e** Ternary phase diagram for three predicted fission-product phases; individual metallic precipitate compositions were derived from clustering analysis of the STEM-EDS dataset.

nanoclusters near the fission-product precipitates shown in Fig. 3c are accompanied by further segregation of the noble metal phases associated with Xe in Fig. 4b, c, indicating a nondiffusional mechanism related to the heterogeneous model of the fission fragments[70,71]. (5) The trapping of nanoscale Xe bubbles at the grain boundaries in Fig. 5 followed by growth into large pores could support the alternative recrystallization theory[72]. (6) The size-dependent resolution related to the segregation of fission

products at the nanoscale could explain the bimodal distribution of Xe bubbles[73]. (7) The decay product, such as Ba, could also occur with the fission fragments and gas, which could be a factor that influences the Xe redistribution. (8) The clustering analysis shown in Fig. 6 illustrates that the variance in the phase composition and structure of the metallic precipitates is related to the fission products and Xe bubbles. (9) The Cs content is relatively low across the radial positions because of its low solubility,

contrary to the enrichment of Cs in the HBS of the MOX fuel[24]. Our results recommend an upgrade in the current modeling process by considering the different types of sinks, including grain boundaries, metallic precipitates, pores, and so on. The Xe mobility at different interfaces must also be incorporated into the current molecular dynamics simulations. Based on all the observations mentioned above, it is plausible to extend this unsupervised ML approach and the cluster analysis to any datasets from SEM-EDS, STEM-EDS, APT, and synchrotron XRD[16–23,25–27,29–31,52,70,74] for nuclear fuel characterization—especially when the datasets are too large or complex to be processed with traditional software or human-centered data exploration. In addition, our approach can be used to improve training sets for supervised ML or DL for any crystallography classification[45] or combined diffraction and chemistry identification, prediction, and merging[75].

The early stage of ML application in materials sciences majorly focused on simple or straightforward algorithms. This study has stepped further to integrate the unsupervised ML into the microscopy characterization of nuclear fuels—one of the challenging problems in materials science field. The nuclear fuel performance heavily depends on the ability of the spent fuel matrix to immobilize the fission products, and the interfacial behavior between fuel and matrix in terms of cracking, fission-product damage, second-phase precipitation, etc. It is apparent that the oxide fuels contain the alkali/transition metals/actinides/lanthanides/noble gases with similar multiphase nanoprecipitates (Fig. 1), where the new chemical compounds generate during nuclear reactions involves absorption, agglomeration, clustering, migration, release, and transport. Current predictive models on these compositionally complex non-equilibrium procedures rely mainly on the mean-field theories. A deeper understanding on this matter with more experimental data would contribute to a more complete and mechanistically based modeling of microstructural and microchemical evolution of heterogeneous nuclear fuels under irradiation. Many fundamental mechanisms related to damage formation, disordering, and amorphization in metal, ceramics, alloy fuels require realistic models that could predict the material's behavior over their full range of length and time scales. Capabilities of our data-driven, ML methods provide a view of the materials with far higher statistical insights than standard X-ray mapping methods are illustrated, and a computationally efficient approach for handling the prodigious amounts of data generated by potential automated mapping shown in Fig. 3. For instance, this data collection pipeline and DL framework from this study could acquire knowledge such as electronic structure from 4D STEM, which would permit the development of realistic interatomic potentials or for novel metals and their oxides and enable large-scale atomistic simulations to provide the underpinning mesoscale models of microstructural evolution. Upgrading the current nuclear fuel characterization experiments with AI would ultimately provide a fundamental basis for developing and predicting the behavior of advanced fuels with the potential for outstanding fuel performance and in turn enable the design of safer and more efficient nuclear energy systems. Moreover, for atomic-scale STEM-EDS maps, future study can include uncertainty quantification for compositional analysis shown in Fig. 6. to dissociate the error in chemical maps from sample preparation[76].

Conclusion. In summary, we developed a high-quality ML-enhanced microscopy method for applications in nuclear fuel analysis. This approach was demonstrated on high-burnup spent nuclear fuels, and it offers novel insights compared to the common notions on obscure nanostructures. Compared with the conventional electron microscopy data collection, this advanced augmented approach accelerates the pace of both post irradiation examination characterization and analysis. It also reveals the detailed nanoclusters with respect to the fission-product precipitates and Xe bubbles associated with the potential onset of the HBS formation. This study highlights the data mining of precipitates chemistry and provides evidence of the ability to extract segregation information from unknown precipitation clusters, addressing the need for precise separation of the elements from analytical spectroscopy for fission products. Based on the insights from the abundance of materials informatics, it is beneficial to leverage such methods to establish a database to reduce uncertainty in fuel-performance modeling, which is also of great importance to the numerous microscopy characterization efforts in interdisciplinary research fields.

Methods

Materials pedigree and sample preparation. The high-burnup fuel specimens were taken from fuel rods extracted from a 9×9 assembly of the Limerick-1 BWR plant and 15×15 assembly of the H.B. Robinson Unit 2 PWR plant. Both fuel rods experienced long-timescale irradiation and reached high-burnup states at low temperatures, which is suitable for the study of HBS. The Limerick fuel rods were irradiated for three cycles. The fuel rod for this study reached an average burnup of approximately 56 MWd kgU^{-1}. The detailed fuel pin design, operation history, and performance data are available in ref. [77]. The H.B. Robinson fuel was from the E02 rod in assembly S-15H and exposed for seven cycles with an average burnup of ~72 MWd kgU^{-1}. The cladding materials are Zircaloy-4 for H.B. Robinson fuels and Zicaloy-2 for Limerick fuels. The irradiation temperature range can be referred to the power history shown in Supplementary Fig. S11. More details regarding fuel pin-relevant characteristics can be found in the paper by Ruzauskas and Fardell[78]. The sample slices were cross-sectioned along the axial direction from the pellets in the fuel rod's central region. The cross-section was fixed in conductive epoxy in a custom aluminum mount and polished inside the hot cell of the Irradiated Fuels Examination Laboratory at Oak Ridge National Laboratory (ORNL). A carbon layer of 5–10 nm was coated on the final sample with a Cressington Scientific Instruments 108 carbon deposition system.

Electron microscopy characterization. A Tescan MIRA3 GMH scanning electron microscope equipped with an Oxford Instruments Symmetry electron backscatter diffraction (EBSD) detector was used for EBSD and TKD scans. All the EBSD scans were conducted using a 20 kV, 4 nA beam with a 0.05-μm step size at a working distance of approximately 12 mm at 70° stage tilt. EBSD camera settings were 2×2 binning (320×240 pixels), a gain setting of 13, 24–28 ms of exposure time, and 64-frame background collection. All the TKD scans were performed using a 20 kV, 2.5 nA beam with a 10-nm step size for better spatial resolution in the submicron restructured grains of the transition and HBS. The acquired EBSD maps were analyzed by the EDAX Orientation Imaging Microscopy (OIM) v8 software and the Aztec 4.3 software package, and indexed using a reference of UO_2 (cubic, space group: Fm3m, No. 225). Site-specific thin foils were lifted out based on the EBSD data using an FEI (now Thermo Fisher Scientific) Quanta DualBeam FIB-SEM system for TKD, TEM, and STEM analysis. The voltage and beam current were gradually reduced during thinning of the TEM lamella, with a final cleaning of 5 keV and 27 pA to minimize the FIB-induced ion beam damage and Ga contamination. In addition, Ga peaks have been deconvolved from the spectral analysis. The TEM analysis utilized a 200 kV FEI Talos F200X S/TEM with the four-detector SuperX EDS system. All the STEM-EDS spectrum-image (SI) maps were recorded using a probe current of ~0.8–1 nA and a sub-nanometer size probe at full-width half-maximum (FWHM) over a 1024×1024 pixel region.

Procedures for the machine-learning analysis. Each raw STEM-EDS SI map was processed by the advanced unsupervised ML methods. The noise-filtered maps generated by the low-rank description via ML are easier to interpret compared with the raw SI data, which unbiasedly identifies small-scale features such as nanoclusters and chemistry segregation[79–82]. The general ML steps include (1) convert noise scale real space to Poisson space[83,84] of the raw SI data, (2) perform SVD and PCA noise-filtering on the chemistry, (3) return to real space, (4) re-orthogonalize, and (5) perform matrix rotations[83–86]. Varimax must be applied to an orthonormal matrix, so the images (spatial simplicity) or spectra (spectral simplicity) must all be magnitude 1 and mutually orthogonal. A simple factor-PCA method is then applied to recover a re-orthogonalized PCA model[83]. Nonnegativity-constrained multivariate curve resolution via the alternating least-squares algorithm[87–90] was also applied to the spatial-simplicity rotations on the optimally scaled PCA solutions. The custom Python script and code flowchart followed the steps of the Keenan method[79,80,83–86]. All the data were reconstructed with 4×4 (2×2 for Xe bubbles) binning in space and ×1 binning in the spectrum between the energy range between 0.2 and 7.2 keV (intergranular Xe bubbles up to 25 keV). The ultimate spatial resolution of the analyzed elemental maps

is typically reduced in comparison to the native resolution of the acquired data. This is highly dependent on the acquisition parameters of the data including electron probe size and quality of the specimen, etc. The use of spatial and spectral binning is often used to improve the local signal-to-noise over the binned region that results in a reduction of spatial resolution and may also lead to a reduction in spectral detail. The selection of the degree of binning that is appropriate for the data is chosen based upon the quality of the data and the information contained within said data. Caution should also be used to avoid analytical artifacts that are not representative of the data that may arise due to over-interpretation/over-analysis. The standardless Cliff–Lorimer method[68] was used to quantify the atomic concentration of the metallic fission-product precipitates. The X-ray peaks were determined via the NMF method[61,62]. Clustering analysis was performed using an SVD[42,91,92] for dimensionality reduction. The k-means clustering algorithm[42,92–95] was applied for identifying precipitate clusters.

Computational methods. The machine-learning algorithms can be categorized into four steps. First, the acquired raw X-ray spectrum maps with the intensity I for each location of pixel (x, y) over the energy bin E. The data cube can be formulated as $I((x, y) | E)$ with a size of total pixels $(x \cdot y) = n$ total energy channels m. This data cube then can be converted to a matrix $D_{n,m}$. This matrix contains counted photon data with Poisson noise. To denoise this noisy data matrix due to the inherent low counting from STEM-EDS, we can convert the D to $\hat{D}$ with scaling factor F and H:[83,96,97]

$$\hat{D} = FDH \quad (1)$$

After applying the scaling step, the heteroscedastic D becomes more homoscedastic $\hat{D}$. F is a diagonal matrix representing the square root of the inverse of the unfolded mean image along the diagonal; H is also a diagonal matrix representing the square root of the inverse of the unfolded mean spectrum along the diagonal. The combined effect is to scale-down high peaks while scale-up low peaks, providing an amenable heteroscedastic variance fitting for the machine-learning analysis. Kotula and Benthem[98] have also defined the boundary condition for extremely low counting rates with the very sparse dataset, where the matrix F can be replaced by constant 1, and Eq. (S1) (full scaling) turns into $\hat{D} = DH$ (spectral-only scaling) to reveal tiny signals from the spectrum images. Then, we could apply step 2 of matrix decomposition with reduced-rank model[99] for matrix $\hat{D}$ such as singular-value decomposition (SVD) and principal component analysis (PCA).

$$\hat{D} \approx ULV^T \quad (2)$$

where matrix U has a size of $n \cdot k$, diagonal matrix L has a size of $k \cdot k$, and matrix V has a size of $m \cdot k$. U and V are mutual orthonormal with unit vector in each column, which indicates they are orthogonal in the high-dimensional space. Each column of U consists of abundance/spatial weighting of each component and each row of transpose of V stores the associated endmember spectrum corresponding to that individual component. That element associates with the associated singular value of the diagonal of L. The PCA procedure can be applied equivalent to the SVD step. Here, we combine the diagonal matrix L into one of the other factors U or V so that the eigenvalues of the principal components can be obtained as the squares of the singular values in the Poisson-noise-scaled space.

$$\hat{D} \approx ULV^T = \hat{A}\hat{S}^T \quad (3)$$

where $\hat{A}$ and $\hat{S}$ denote the scores and loadings of the PCA, respectively. Depending on the ambiguity of the intensity from the factor-analysis, either $UL = \hat{A}$ and $V^T = \hat{S}^T$ or $U = \hat{A}$ and $LV^T = \hat{S}^T$. Next, we need find the inverse of matrix $\hat{A}$ and $\hat{S}$. Considering the singular matrix where orthogonality and orthonormality can be broken, a factor-PCA (fPCA) algorithm is implemented into our Python code, where a re-orthogonalized factor model has been generated[55,85,86].

$$D \approx AS^T = TP^T \quad (4)$$

Here, a new matrix D has been obtained as the "re-orthogonalized solution", where T and P are re-orthogonalized basis. The last step is to examine the validity of the new model, so-called Varimax criterion:[100–102]

$$D \approx (TR)(R^{-1}P)^T = \widetilde{T}\widetilde{P}^T \quad (5)$$

where R is a non-singular matrix. Varimax can simplifies the data interpretation so that an orthogonal rotation $R^T = R^{-1}$ can be found for matrix R. At a given pixel or a single energy channel P, matrix R can rotate the abundance T or the endmember P into a single component representing a strong value. Such simplicity method can provide a relative significance measurement of each component through a combination of diagonal elements $\left[(P^T P)(T^T T) \right]$ or $\left[(\widetilde{P}^T \widetilde{P})(\widetilde{T}^T \widetilde{T}) \right]$[103].

Regarding the data mining and informatics in Fig. 6, the clustering algorithms are performed on the spectral-only data from the standardless Cliff–Lorimer method[68]. First, the The k-means clustering algorithm divides a set of compositional quantifications N samples X into K disjoint clusters C; the mean variance known as inertia or within-cluster sum-of-squares can be described as the minimization of inertia criterion below[104]:

$$\sum_{i=0}^{n} \min_{\mu_j \in C} \left(||x_i - \mu_j||^2 \right) \quad (6)$$

The means are defined by cluster centroids in our large datasets of compositional quantifications. This algorithm requires the number of clusters to be specified. The selection of clusters can be referred to the elbow curve shown in Supplementary Fig. S1. The followed-up PCA aims to reduce the dimensionality in the case of very high-dimensional spaces where the Euclidean distances inflates. An example of a sparse PCA method has been given here to solve the noisy STEM-EDS quantification problem. For a given matrix X contains n rows of components in R_p, the decomposition coefficient can be described through a linear combination as matrix U and V. This decomposition procedure via PCA solution can then be represented with an l_1 penalty on the components[105–107]:

$$(U^*, V^*) = \underset{U,V}{\operatorname{argmin}} \, 0.5||X - UV||^2_{Fro} + \alpha ||V||_{1,1} : \text{subject to} ||U_k||_2 \leq 1 \text{ for all } 0 \leq k \leq n \quad (7)$$

where $||\cdot||_{Fro}$ and $||\cdot||_{Fro}$ stand for the Frobenius norm stands and the entry-wise matrix norm (the sum of the absolute values of all the entries in the matrix), respectively.

Data availability

Additional data and further information are available from the corresponding authors on request.

Code availability

The computer code used in this study can be found in the GitHub repository (https://github.com/keyoumao/ML_FUEL_CM_COMMSMAT).

References

1. Ewing, R. C. Long-term storage of spent nuclear fuel. *Nat. Mater.* **14**, 252–257 (2015).
2. Hopp, T., Zok, D., Kleine, T. & Steinhauser, G. Non-natural ruthenium isotope ratios of the undeclared 2017 atmospheric release consistent with civilian nuclear activities. *Nat. Commun.* **11**, 1–7 (2020).
3. Markard, J. The next phase of the energy transition and its implications for research and policy. *Nature Energy* **3**, 628–633 (2018).
4. Kramer, G. J. & Haigh, M. No quick switch to low-carbon energy. *Nature* **462**, 568–569 (2009).
5. Grimes, R. W. & Nuttall, W. J. Generating the option of a two-stage nuclear renaissance. *Science* **329**, 799–803 (2010).
6. Wilson, I. A. G. & Staffell, I. Rapid fuel switching from coal to natural gas through effective carbon pricing. *Nat. Energy* **3**, 365–372 (2018).
7. Konings, R. J. M., Wiss, T. & Beneš, O. Predicting material release during a nuclear reactor accident. *Nat. Mater.* **14**, 247–252 (2015).
8. Adachi, T., Muromura, T., Takeishi, H. & Yamamoto, T. Metallic phases precipitated in UO2 fuel. II. Insoluble residue in simulated fuel. *J. Nucl. Mater.* **160**, 81–87 (1988).
9. Muromura, T. et al. Metallic phases precipitated in UO2 fuel. I. Phases in simulated fuel. *J. Nucl. Mater.* **151**, 327–333 (1988).
10. Tonks, M. R. et al. Mechanistic materials modeling for nuclear fuel performance. *Ann. Nucl. Energy* **105**, 11–24 (2017).
11. Bai, X. M., Tonks, M. R., Zhang, Y. & Hales, J. D. Multiscale modeling of thermal conductivity of high burnup structures in UO2 fuels. *J. Nucl. Mater.* **470**, 208–215 (2016).
12. Teague, M. C., Fromm, B. S., Tonks, M. R. & Field, D. P. Using coupled mesoscale experiments and simulations to investigate high burn-up oxide fuel thermal conductivity. *Jom* **66**, 2569–2577 (2014).
13. Tonks, M. R. et al. Development of a multiscale thermal conductivity model for fission gas in UO2. *J. Nucl. Mater.* **469**, 89–98 (2016).
14. Permann, C. J. et al. MOOSE: enabling massively parallel multiphysics simulation. *SoftwareX* **11**, 100430 (2020).
15. Yip, S. & Short, M. P. Multiscale materials modelling at the mesoscale. *Nat. Mater.* **12**, 774–777 (2013).
16. Harp, J. M., Porter, D. L., Miller, B. D., Trowbridge, T. L. & Carmack, W. J. Scanning electron microscopy examination of a fast flux test facility irradiated U-10Zr fuel cross section clad with HT-9. *J. Nucl. Mater.* **494**, 227–239 (2017).
17. Harp, J. M., Capriotti, L., Porter, D. L. & Cole, J. I. U-10Zr and U-5Fs: fuel/cladding chemical interaction behavior differences. *J. Nucl. Mater.* **528**, 151840 (2020).

18. Harp, J. M., Capriotti, L. & Chichester, H. J. M. Postirradiation examination of FUTURIX-FTA metallic alloy experiments. *J. Nucl. Mater.* **515**, 420–433 (2019).

19. Parrish, R. J., Cappia, F. & Aitkaliyeva, A. Comparison of the radial effects of burnup on fast reactor MOX fuel microstructure and solid fission products. *J. Nucl. Mater.* **531**, 152003 (2020).

20. Cappia, F. et al. Electron microscopy characterization of fast reactor MOX joint oxyde-gaine (JOG). *J. Nucl. Mater.* **531**, 151964 (2020).

21. Bouloré, A., Aufore, L., Federici, E., Blanpain, P. & Blachier, R. Advanced characterization of MIMAS MOX fuel microstructure to quantify the HBS formation. *Nucl. Eng. Des.* **281**, 79–87 (2015).

22. Morimoto, K. et al. Preparation and characterization of (Pu, U, Np, Am, simulated FP) O 2-X. *J. Phys. Chemistry Solids* **66**, 634–638 (2005).

23. Gerczak, T. J., Parish, C. M., Edmondson, P. D., Baldwin, C. A. & Terrani, K. A. Restructuring in high burnup UO₂ studied using modern electron microscopy. *J. Nucl. Mater.* **509**, 245–259 (2018).

24. Noirot, J. et al. Heterogeneous UO2 fuel irradiated up to a high burn-up: investigation of the HBS and of fission product releases. *J. Nucl. Mater.* **442**, 309–319 (2013).

25. Clark, R. A. et al. Distribution of metallic fission-product particles in the cladding liner of spent nuclear fuel. *npj Mater. Degrad* **4**, 4 (2020).

26. Wen, H., van Rooyen, I. J., Hunn, J. D. & Gerczak, T. J. Electron microscopy study of Pd, Ag, and Cs in carbon areas in the locally corroded SiC layer in a neutron-irradiated TRISO fuel particle. *J. Eur. Ceram. Soc.* **38**, 4173–4188 (2018).

27. Meher, S., van Rooyen, I. J. & Jiang, C. Understanding of fission products transport in SiC layer of TRISO fuels by nanoscale characterization and modeling. *J. Nucl. Mater.* **527**, 151793 (2019).

28. Bargmann, S. et al. Generation of 3D representative volume elements for heterogeneous materials: a review. *Prog. Mater. Sci.* **96**, 322–384 (2018).

29. Bachhav, M. et al. A novel approach to determine the local burnup in irradiated fuels using atom probe tomography (APT). *J. Nucl. Mater.* **528**, 151853 (2020).

30. Kautz, E., Burkes, D., Joshi, V., Lavender, C. & Devaraj, A. Nanoscale spatially resolved mapping of uranium enrichment. *Sci. Rep.* **9**, 12302 (2019).

31. Thomas, J. et al. The application of synchrotron micro-computed tomography to characterize the three-dimensional microstructure in irradiated nuclear fuel. *J. Nucl. Mater.* **537**, 152161 (2020).

32. Wood, V. X-ray tomography for battery research and development. *Nat. Rev. Mater* **3**, 293–295 (2018).

33. Goldstein, J. I. et al. Scanning electron microscopy and x-ray microanalysis. *Scanning Electron Microsc. X-ray Microanal.* 1–550. https://doi.org/10.1007/978-1-4939-6676-9 (2017).

34. Lugg, N. R., Kothleitner, G., Shibata, N. & Ikuhara, Y. On the quantitativeness of EDS STEM. *Ultramicroscopy* **151**, 150–159 (2015).

35. Goldstein, J. I. et al. Energy dispersive X-ray spectrometry: physical principles and user-selected parameters. *Scanning Electron Microsc. X-Ray Microanal.* 209–234. https://doi.org/10.1007/978-1-4939-6676-9_16 (2018).

36. Spurgeon, S. R. et al. Towards data-driven next-generation transmission electron microscopy. *Nat. Mater.* https://doi.org/10.1038/s41563-020-00833-z (2020).

37. Kim, H. K. et al. Nanoscale light element identification using machine learning aided STEM-EDS. *Sci. Rep.* **10**, 1–12 (2020).

38. Han, Y. et al. Deep learning STEM-EDX tomography of nanocrystals. *Nat. Mach. Intell* **3**, 267–274 (2021).

39. Chen, H., Nabiei, F., Badro, J., Alexander, D. T. & Hébert, C. Machine learning on STEM-EDS data for quantifying overlapping deep-mantle rock assemblages. *Microsc. Microanal.* **26**, 1878–1880 (2020).

40. Ophus, C. Four-dimensional scanning transmission electron microscopy (4D-STEM): from scanning nanodiffraction to ptychography and beyond. *Microsc. Microanal.* **25**, 563–582 (2019).

41. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vols 2016-Decem. (IEEE, 2016).

42. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

43. Abadi, M. et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* preprint arXiv:1603.04467 (2016).

44. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. *Adv. Neural Inform. Processing Sys.* **32**, 1–12 (2019).

45. Aguiar, J. A., Gong, M. L., Unocic, R. R., Tasdizen, T. & Miller, B. D. Decoding crystallography from high-resolution electron imaging and diffraction datasets with deep learning. *Sci. Adv.* **5**, eaaw1949 (2019).

46. Villars P. & Cenzual K. *Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds. Choice Reviews Online*, Vol. 45 (ASM International, 2008).

47. Vaitkus, A., Merkys, A. & Grazulis, S. Validation of the crystallography open database using the crystallographic information framework. *J. Appl. Crystallogr.* **54**, 661–672 (2021).

48. Jain, A. et al. The materials project: accelerating materials design through theory-driven data and tools. *Handb. Mater. Model.* 1–34. https://doi.org/10.1007/978-3-319-42913-7_60-1 (2018).

49. Matzke, H. & Spino, J. Formation of the rim structure in high burnup fuel. *J. Nucl. Mater.* **248**, 170–179 (1997).

50. Romano, A., Horvath, M. I. & Restani, R. Evolution of porosity in the high-burnup fuel structure. *J. Nucl. Mater.* **361**, 62–68 (2007).

51. Lassmann, K., Walker, C. T., van de Laar, J. & Lindström, F. Modelling the high burnup UO2 structure in LWR fuel. *J. Nucl. Mater.* **226**, 1–8 (1995).

52. Rondinella, V. V. & Wiss, T. The high burn-up structure in nuclear fuel. *Materials Today* **13**, 24–32 (2010).

53. Bramman, J. I., Sharpe, R. M., Thom, D. & Yates, G. Metallic fission-product inclusions in irradiated oxide fuels. *J. Nucl. Mater.* **25**, 201–215 (1968).

54. Kleykamp, H., Paschoal, J. O., Pejsa, R. & Thümmler, F. Composition and structure of fission product precipitates in irradiated oxide fuels: correlation with phase studies in the Mo-Ru-Rh-Pd and BaO-UO2-ZrO2-MoO2 systems. *J. Nucl. Mater.* **130**, 426–433 (1985).

55. Keenan, M. R. Multivariate analysis of spectral images composed of count data. In *Techniques and Applications of Hyperspectral Image Analysis* (Eds Grahn, H. & Geladi, P.) 89–126 (John Wiley & Sons, Ltd, 2007).

56. Moen, E. et al. Deep learning for cellular image analysis. *Nat. Methods* **16**, 1233–1246 (2019).

57. von Chamier, L. et al. Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* **12**, 1–18 (2021).

58. Sosa, J. M., Huber, D. E., Welk, B. & Fraser, H. L. Development and application of MIPAR™: a novel software package for two- and three-dimensional microstructural characterization. *Integr. Mater. Manuf. Innov* **3**, 123–140 (2014).

59. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).

60. Weber, K. et al. RGB marking facilitates multicolor clonal cell tracking. *Nat. Med.* **17**, 504–509 (2011).

61. Kannan, R. et al. Deep data analysis via physically constrained linear unmixing: universal framework, domain examples, and a community-wide platform. *Adv. Struct. Chem. Imaging* **4**, 6 (2018).

62. Stanev, V. et al. Unsupervised phase mapping of X-ray diffraction data by nonnegative matrix factorization integrated with custom clustering. *npj Comput. Mater.* **4**, 43 (2018).

63. Cooper, M. W. D., Stanek, C. R. & Andersson, D. A. The role of dopant charge state on defect chemistry and grain growth of doped UO₂. *Acta Mater.* **150**, 403–413 (2018).

64. Aagesen, L. K., Schwen, D., Tonks, M. R. & Zhang, Y. Phase-field modeling of fission gas bubble growth on grain boundaries and triple junctions in UO₂ nuclear fuel. *Comput. Mater. Sci.* **161**, 35–45 (2019).

65. Matthews, C., Perriot, R., Cooper, M. W. D., Stanek, C. R. & Andersson, D. A. Cluster dynamics simulation of xenon diffusion during irradiation in UO₂. *J. Nucl. Mater.* **540**, 152326 (2020).

66. Jaumot, J., de Juan, A. & Tauler, R. MCR-ALS GUI 2.0: new features and applications. *Chemom. Intell. Lab. Syst.* **140**, 1–12 (2015).

67. Stork, C. L. & Keenan, M. R. Advantages of clustering in the phase classification of hyperspectral materials images. *Microsc. Microanal.* **16**, 810–820 (2010).

68. Cliff, G. & Lorimer, G. W. The quantitative analysis of thin specimens. *J. Microsc.* **103**, 203–207 (1975).

69. Spino, J. & Papaioannou, D. Lattice parameter changes associated with the rim-structure formation in high burn-up UO2 fuels by micro X-ray diffraction. *J. Nucl. Mater.* **281**, 146–162 (2000).

70. Rest, J. et al. Fission gas release from UO₂ nuclear fuel: a review. *J. Nucl. Mater.* **513**, 310–345 (2019).

71. Tonks, M. et al. Unit mechanisms of fission gas release: current understanding and future needs. *J. Nucl. Mater.* **504**, 300–317 (2018).

72. Rest, J. & Hofman, G. L. Alternative explanation for evidence that xenon depletion, pore formation, and grain subdivision begin at different local burnups. *J. Nucl. Mater.* **277**, 231–238 (2000).

73. Govers, K. et al. Molecular dynamics study of Xe bubble re-solution in UO₂. *J. Nucl. Mater.* **420**, 282–290 (2012).

74. Nogita, K. & Une, K. High resolution TEM observation and density estimation of Xe bubbles in high burnup UO₂ fuels. *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms* **141**, 481–486 (1998).

75. Aguiar, J. A., Gong, M. L. & Tasdizen, T. Crystallographic prediction from diffraction and chemistry data for higher throughput classification using machine learning. *Comput. Mater. Sci.* **173**, 109409 (2020).

76. Spurgeon, S. R., Du, Y. & Chambers, S. A. Measurement error in atomic-scale scanning transmission electron microscopy-energy-dispersive X-ray spectroscopy (STEM-EDS) mapping of a model oxide interface. *Microsc. Microanal* **23**, 513–517 (2017).

77. EPRI TR-1003222. *Poolside Examination Results and Assessment, GE11 BWR Fuel Exposed to 52 to 65 GWd/MTU at the Limerick 1 and 2 Reactors.* https://www.epri.com/research/products/000000000001003222 (2002).

78. Ruzauskas, E. J. & Fardell, K. N. *Design, operation, and performance data for high burnup PWR fuel from the HB Robinson plant for use in the NRC experimental program at Argonne National Laboratory.* Electric Power Research Institute, EPRI Report 1001558 (2001).

79. Edmondson, P. D., Parish, C. M. & Nanstad, R. K. Using complimentary microscopy methods to examine Ni-Mn-Si-precipitates in highly-irradiated reactor pressure vessel steels. *Acta Mater.* **134**, 31–39 (2017).

80. Parish, C. M. & Brewer, L. N. Multivariate statistics applications in phase analysis of STEM-EDS spectrum images. *Ultramicroscopy* **110**, 134–143 (2010).

81. Parish, C. M. & Miller, M. K. A review of advantages of high-efficiency X-ray spectrum imaging for analysis of nanostructured ferritic alloys. *J. Nucl. Mater.* **462**, 433–442 (2015).

82. Parish, C. M. & Miller, M. K. Aberration-corrected X-ray spectrum imaging and fresnel contrast to differentiate nanoclusters and cavities in helium-irradiated alloy 14YWT. *Microsc. Microanal.* **20**, 613–626 (2014).

83. Grahn, H. F. & Geladi, P. Techniques and applications of hyperspectral image analysis. *Tech. Appl. Hyperspectral Image Anal.* 1–368. https://doi.org/10.1002/9780470010884 (2007).

84. Parish, C. M. Multivariate statistics applications in scanning transmission electron microscopy X-ray spectrum imaging. In *Advances in Imaging and Electron Physics* Vol. 168, (ed. Hawkes, P. W.) 249–295 (Academic Press Inc., 2011).

85. Kotula, P. G. & Keenan, M. R. Application of multivariate statistical analysis to STEM X-ray spectral images: interfacial analysis in microelectronics. *Micros. Microanal.* **12**, 538–544 (2006).

86. Kotula, P. G., Keenan, M. R. & Michael, J. R. Automated analysis of SEM X-ray spectral images: a powerful new microanalysis tool. *Microsc. Microanal.* **9**, 1–17 (2003).

87. Jones, H. D. T. et al. Weighting hyperspectral image data for improved multivariate curve resolution results. *J. Chemom.* **22**, 482–490 (2008).

88. Tauler, R., Smilde, A. & Kowalski, B. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.* **9**, 31–58 (1995).

89. Jaumot, J., Gargallo, R., De Juan, A. & Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom. Intell. Lab. Syst.* **76**, 101–110 (2005).

90. de Juan, A. & Tauler, R. Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. *Crit. Rev. Anal. Chem.* **36**, 163–176 (2006).

91. Tipping, M. E. & Bishop, C. M. Mixtures of probabilistic principal component analyzers. *Neural Comput* **11**, 443–482 (1999).

92. Jollife, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci* **374**, 20150202 (2016).

93. Carvalho, M. J., Melo-Gonçalves, P., Teixeira, J. C. & Rocha, A. Regionalization of Europe based on a K-means cluster analysis of the climate change of temperatures and precipitation. *Phys. Chem. Earth* **94**, 22–28 (2016).

94. Cohen, M. B., Elder, S., Musco, C., Musco, C. & Persu, M. Dimensionality reduction for k-means clustering and low rank approximation. *Proc. Annu. ACM Symp. Theory Comput.* **14-17-June**, 163–172 (2015).

95. Mokdad, F. & Haddad, B. Improved infrared precipitation estimation approaches based on k-means clustering: application to north Algeria using MSG-SEVIRI satellite data. *Adv. Sp. Res.* **59**, 2880–2900 (2017).

96. Keenan, M. R. & Kotula, P. G. Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. *Surf. Interface Anal.* **36**, 203–212 (2004).

97. Keenan, M. R. & Kotula, P. G. Optimal scaling of TOF-SIMS spectrum-images prior to multivariate statistical analysis. *Appl. Surf. Sci.* **231–232**, 240–244 (2004).

98. Kotula, P. G. & Benthem Van, M. H. Revisiting noise scaling for multivariate statistical analysis. *Microsc. Microanal.* **21**, 1423–1424 (2015).

99. Stewart, G. W. On the early history of the singular value decomposition. *SIAM Rev.* **35**, 551–566 (1993).

100. Keenan, M. R. Exploiting spatial-domain simplicity in spectral image analysis. *Surf. Interface Anal.* **41**, 79–87 (2009).

101. Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**, 187–200 (1958).

102. Sherin, R. J. A matrix formulation of Kaiser's varimax criterion. *Psychometrika* **31**, 535–538 (1966).

103. Smentkowski, V. S., Ostrowski, S. G. & Keenan, M. R. A comparison of multivariate statistical analysis protocols for ToF-SIMS spectral images. *Surf. Interface Anal.* **41**, 88–96 (2009).

104. Hamerly, G. & Elkan, C. Learning the k in kmeans. *Adv. neural Inf. Process.* **17**, 1–8 (2004).

105. Jenatton, R., Obozinski, G. & Bach, F. Structured Sparse Principal Component Analysis. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics.* In Proceedings of Machine Learning Research 366–373, Available from https://proceedings.mlr.press/v9/jenatton10a.html (2010).

106. Mairal, J., Bach, F., Ponce, J. & Sapiro, G. Online dictionary learning for sparse coding. *ACM Int. Conf. Proceeding Ser.* **382**, 689–696 (2009).

107. Szlam, A., Kluger, Y. & Tygert, M. An implementation of a randomized algorithm for principal component analysis. *arXiv* preprint arXiv:1412.3510. 1–13 (2014).

Acknowledgements

Author contributions

K.S.M. conducted the experiments and machine-learning analysis and drafted the manuscript; T.J.G., J.M.H., and K.A.T. provided the high-value fuels, oversaw general execution, coordinated the sample transfer, and contributed to the theoretical explanations of the experimental data. C.S.M. aided in sample preparation, data collections, and discussions; T.G.L. and O.K. aided in the microscopy data collection and analysis; A.T.N. provided conceptualization, funding acquisition, and regular guidance. C.M.P. and P.D.E. performed the machine-learning analysis, provided critical interpretation, and review, and P.D.E. finalized the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43246-022-00244-4.

Correspondence and requests for materials should be addressed to Keyou S. Mao, Tyler J. Gerczak, Chad M. Parish or Philip D. Edmondson.

Peer review information *Communications Materials* thanks Arun Devaraj, Brian DeCost and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editor: John Plummer. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.