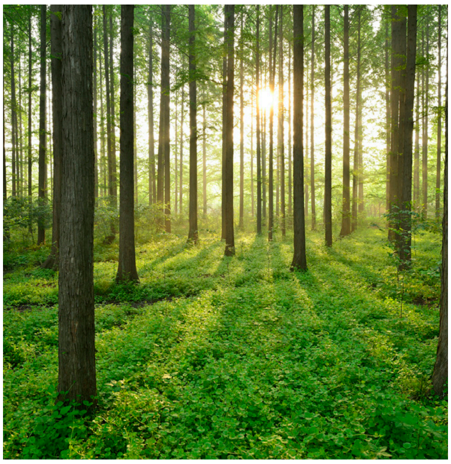
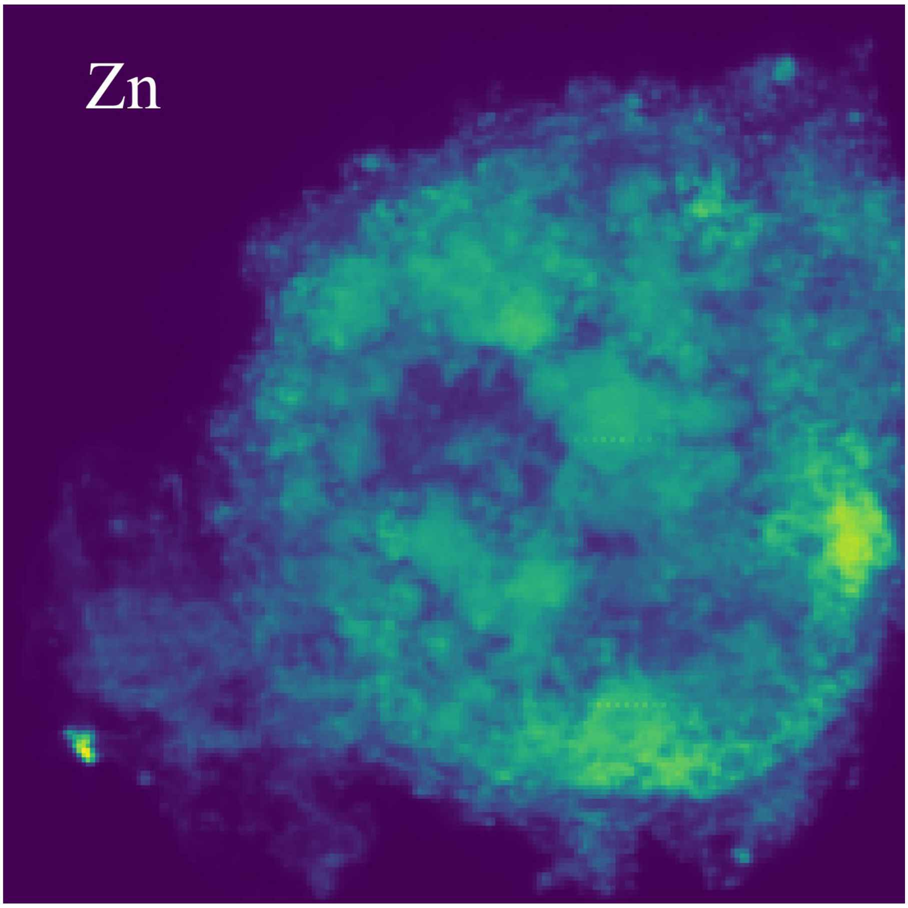
NANO-XRF ANALYSIS FOR METAL SPECIATION IN ASH

REPORT 2026:1187



THE ASH PROGRAMME



Energiforsk

Nano-XRF Analysis for Metal Speciation in Ash

Evaluation of data analysis methods

FANNY BERGMAN, JENNY RISSLER, LINDA HARTMAN, KARIN KARLFELDT FEDJE

ISBN 978-91-89917-30-9 | © Energiforsk April 2026

Energiforsk AB | Phone: 08-677 25 30 | E-mail: kontakt@energiforsk.se | www.energiforsk.se

Foreword

It is a challenge to identify which metals are present in fly ash. This is mainly because the ash contains many elements in various chemical forms. The chemical form in which the metals are present is crucial to understanding how they can leach into the environment and what possibilities there are for chemically recovering the metals.

Nano-resolved X-ray fluorescence (nXRF) microscopy provides a possibility to retrieve information about chemical form of metals, via analysis of spatial colocalization of different elements.

In this project, Fanny Bergman, Jenny Rissler, Linda Hartman and Karin Karlfeldt Fedje have identified and evaluated statistical methods for data analysis of spatially (in 2D) resolved elemental data, with the aim of identifying elements often colocalized with Zn, a metal of importance both from the environmental leaching and extraction perspective.

Stockholm in April 2026

Marie Kofod-Hansen
Programme manager, Ash programme

These are the results and conclusions of a project, which is part of a research programme run by Energiforsk. The author/authors are responsible for the content.

Förord

Det är en utmaning att karlägga vilka metaller som förekommer i flygaskor. Det beror främst på att askan innehåller många grundämnen i flera kemiska former. Vilken kemisk fom metaller förekommer i är avgörande för att förstå hur den kan laka till miljön och vilka möjligheter det finns att kemiskt återvinna metallerna.

Nanoupplöst röntgenfluorescensmikroskopi (nXRF) är en teknik som kan användas för att få information om till exempel kemiska former hos metaller i askan genom att analysera hur olika grundämnen ligger samlokaliserade i askpartiklar.

I det här projektet har Fanny Bergman, Jenny Rissler, Linda Hartman och Karin Karlfeldt Fedje identifierat och utvärderat statistiska metoder för analys av grundämneskartor (2D) för att identifiera vilka grundämnen som ofta förekommer tillsammans med Zn, en metall som är av betydelse både från ett återvinningsperspektiv, men som också medför miljörisker vid lakning.

Stockholm i April 2026

Marie Kofod-Hansen
Programansvarig Askprogrammet

Här redovisas resultat och slutsatser från ett projekt inom ett forskningsprogram som drivs av Energiforsk. Det är rapportförfattaren/-författarna som ansvarar för innehållet.

Summary

Metals in waste incineration fly ash pose a challenge as their presence often leads to the ash being classified as hazardous waste. Yet these metals also represent valuable resources for recycling. Recovering metals from ash reduces the need for mining, improving resource utilization, and lowers the risk of metals leaching into the environment. Furthermore, the ash/ash residues with low or non-soluble content of metals can be used in construction materials. The chemical forms of metals in the ashes are essential, as they govern their solubility. Thus, knowledge of metal solubility is critical both in minimizing environmental leaching and for improving metal extraction processes.

Determining the chemical form of metals in ash is a challenge, mainly because the ash contains numerous elements, in several chemical forms, and includes both crystalline and amorphous states. In addition, many of the metals of interest, e.g. Cu, Zn, and Pb, occur in comparatively low concentrations, usually around a few weight percent and lower.

Nano-resolved X-ray fluorescence (nXRF) microscopy provides a possibility to retrieve information about chemical form of metals, via analysis of spatial colocalization of different elements. While nXRF provides excellent spatial resolution and low detection limits for many metals, it is important to stress that nXRF cannot distinguish chemical binding from the mere vicinity of elements. nXRF mapping typically generates large data sets which can be challenging to process and interpret as a whole.

In this project and report, we identify and evaluate statistical methods for data analysis of spatially (in 2D) resolved elemental data, with the aim of identifying elements often colocalized with Zn, a metal of importance both from the environmental leaching and extraction perspective. The focus is here on nXRF data, but the methods could in principle also be used for other data types such as scanning electron microscopy energy dispersive spectroscopy (SEM-EDS) maps. Our long-term objective is to identify and propose chemical forms of Zn in fly ash from waste incineration. These findings will be the focus of a subsequent open-access scientific paper, while the present report focus primarily on identifying and evaluating different statistical models for the purpose.

The methods that were identified and evaluated are cluster analysis, principal component analysis (PCA), partial least square regression (PLSR), as well as more manual evaluation methods in combination with linear regression.

We found that manual methods, here exemplified as visual selection of pixels (data points) with certain elemental correlation trends, had the key advantage over the more complex mathematical models that interpretation is relatively straight forward. On the other hand, manual evaluation approaches may be biased, focusing on analyzing clear trends and overlooking complex associations, and does probably not utilize all collected data.

The clustering algorithm identified as relevant and evaluated, HDBSCAN, was the model that showed most realistic results in the sense that it resulted in clusters with physiochemically meaningful interpretations. This approach could categorize less than half of the Zn-containing pixels into a cluster, meaning that a lot of the information in the data set was not used. A likely explanation for the model's difficulty in classifying many of the pixels is that ash particles contain mixtures of Zn species, where mixing ratios vary across pixels. Clustering is also challenging when Zn is present in solid solutions with varying degrees of substitution. However, when the model succeeds in identifying clusters, it may provide useful information by showing which elements are associated with Zn within the identified clusters and by indicating that particles with a certain elemental composition often occur in less complex mixes. It should be kept in mind though, that with this approach the major Zn species may be overlooked if typically occurring in complex and varying mixtures with other forms.

Multivariate latent-variable models were also identified as important, where-of PCA and PLSR were evaluated. Latent variable models can hypothetically be used to identify chemical forms present in the ash, as it reduces the data dimensionality, highlighting underlying chemical patterns. Best case scenario, the extracted components represent distinct chemical species and can be combined to represent mixed phases.

Our evaluation of PCA for the nXRF data show that PCA can be useful to evaluate trends and major chemical form present in the ash. For the purpose of identifying the chemical forms of metals in low content, such as Zn, PCA has the disadvantage that it does not allow specifying a target element. A potential approach could be to combine PCA (or similar) with subsequent analysis of Zn in identified components.

Using PLSR, it is possible to model specific outcome variables, here Zn. While PLSR could suggest patterns explaining some of the Zn variation, no obvious Zn chemical forms could be suggested from the model.

A major challenge interpreting mathematical models such as PLSR and PCA is translating the results into physical or chemical explanations. In the case of both PCA and PLSR, the models often suggest negative weights of elements, which are not readily translatable as physicochemical properties.

We therefore suggest that future analytical efforts include methods such as non-negative matrix factorization (NMF) or positive matrix factorization (PMF), which yield only non-negative weights. Although we here identify these methods as potentially suitable, we have not evaluated them for the nXRF dataset within the current project. We recommend that their applicability, as well as careful consideration on incorporation of spatial information, should be explored in subsequent work.

Based on the evaluation of a test dataset using the different statistical methods identified during the project, three main particle types could be distinguished:

1. Particles with a complex composition, containing nearly all analyzed elements. These are likely melts, originating from material entrained through the combustion bed.

2. Salt particles, which are likely formed through condensation, in contrast to particles associated with entrainment.
3. Calcium-dominated particles, which according to the cluster analysis occurs as a separate group and are not part of the melt fractions (#1).

Zinc was present to some extent in all particle types, although the third type contained the lowest Zn levels. However, these findings are based on data that was analyzed with the method evaluation as the primary purpose. The analysis therefore needs to be refined and made more thoroughly before robust scientific conclusions can be drawn.

Altho not evaluated within this we identified

Keywords

Fly ash, nano resolved XRF (Xray fluorescence spectroscopy), methodology for analysing data, metals, circularity, WtE

Flygaska, nano-upplöst XRF (röntgenfluorescensspektroskopi), analysmetodik, metaller, cirkuläritet, WtE

Sammanfattning

Förekomsten av metaller i flygaska är en miljömässig utmaning, men öppnar också en möjlighet för materialåtervinning. Att återvinna metall ur flygaska bidrar till minskad gruvbrytning, effektivare resursanvändning och minskad risk för lakning av metaller till miljön från askrester. Askan kan även användas som konstruktionsmaterial, förutsatt att lakning av metaller är låg. Vilken kemisk form metaller förekommer i är avgörande för deras löslighet och därmed viktig både för att förstå lakning till miljön och för att utveckla effektivare extraktionsprocesser för kemisk materialåtervinning.

Att bestämma den kemiska formen av metaller i aska är dock en utmaning, främst eftersom askan innehåller många grundämnen i flera kemiska former, inklusive amorfa föreningar. Dessutom förekommer många av de relevanta metallerna (så som Cu, Zn och Pb) i förhållandevis låga koncentrationer, vanligtvis runt några procent eller lägre.

Nanoupplöst röntgenfluorescensmikroskopi (nXRF) är en teknik som kan användas för att få information om t.ex. kemiska former hos metaller i askan genom att analysera hur olika grundämnen ligger samlokaliserade i askpartiklar. nXRF ger hög spatial upplösning (~70 nm) och har låga detektionsgränser för många metaller, men kan egentligen inte skilja kemisk bindning från enbart närhet till ett annat grundämne. Däremot kan en samlokalisering av ämnen vara en indikation på att de är kemiskt bundna. Högupplöst nXRF genererar stora datamängder, vilka kan vara utmanande och resurskrävande att analysera och tolka i sin helhet.

Inom detta projekt identifierar och utvärderar vi statistiska metoder för analys av grundämneskartor (2D) för att identifiera vilka grundämnen som ofta förekommer tillsammans med Zn, en metall som är av betydelse både från ett återvinningsperspektiv, men som också medför miljörisker vid lakning. För vår utvärdering använder vi nXRF-data, men metoderna skulle i princip också kunna användas för andra datatyper såsom data från SEM-EDS. Vårt övergripande mål är att identifiera och föreslå möjliga kemiska former av Zn i flygaska från avfallsförbränning. Slutresultaten av detta, inklusive detaljer om asksammansättningen, planerar vi att publicera i en öppen (fritt tillgänglig) vetenskaplig artikel. Denna rapport fokuserar därför främst på identifiering och utvärdering av utvalda statistiska analysmetoder.

De metoder som identifierades och utvärderades omfattar klusteranalys, principal component analysis (PCA), partial least square regression (PLSR) samt en mer manuell metod där en delmängd av data med intressanta trender selekteras och analyseras med linjärregression.

Manuella utvärderingsmetoder, så som selektering av pixlar (datapunkter) som följer en viss korrelationstrend, har i jämförelse med de mer matematiskt avancerade metoderna fördelen att vara relativt rättfram att tolka. Å andra sidan kan manuella metoder riskera att vara subjektiva i den bemärkelsen att enkla och tydliga trender premieras framför mer komplexa samband, vilka då potentiellt sett skulle missas. Dessutom är det tidskrävande att genom manuell utvärdering utnyttja alla insamlade datapunkter.

Den utvärderade klusteranalysmetoden, HDBSCAN, var den statistiska metod som gav mest realistiska och användbara resultat utifrån kriteriet att ge kemiskt meningsfulla grupperingar av grundämnen. Dock kunde mindre än hälften av alla Zn-innehållande pixlar kategoriseras till ett kluster. Att modellen hade svårt att klassificera många av pixlarna beror troligen på att askpartiklarna innehåller blandningar av olika Zn-former, där blandningsförhållandet skiljer sig åt, snarare än på metoden i sig. Klustring är också utmanande när Zn finns i form av fasta lösningar med varierande substitutionsgrad. I de fall när modellen lyckas identifiera kluster så kan den bidra med användbar information om vilka grundämnen som är associerade med Zn i respektive kluster, samt ge oss information om att dessa Zn-former, åtminstone till viss del, förekommer i icke-blandade former. Det är dock viktigt att tänka på att huvudsakliga Zn-former skulle kunna missas med denna metod, om de förekommer i komplexa och varierande blandningar.

För partiklar med mer komplexa blandningar identifierades multivariata modeller som viktiga. Av dessa valdes PCA och PLSR ut för utvärdering. Dessa modeller kan i teorin användas för att identifiera underliggande mönster och möjliga kemiska former i askan trots att de olika formerna finns representerade i varierande andel i olika pixlar.

Vår utvärdering visade att PCA kan vara ett användbart verktyg för att identifiera de dominerande kemiska föreningarna som förekommer i flygaska och visa på generella trender. Men för att identifiera den kemiska formen av metaller som förekommer i låga koncentrationer, så som Zn, har PCA nackdelen att analysen inte går att styra mot ett specifikt grundämne. Det innebär att de huvudkomponenter man får fram med en PCA inte nödvändigtvis innehåller speciellt mycket Zn. Ett alternativ är då att kombinera PCA - eller liknande modeller - med ett efterföljande analyssteg fokuserat på Zn i de identifierade komponenterna.

Med PLSR är det möjligt att specificera utfallsvariabler (som Zn). Baserat på PLSR kunde mönster som förklarade en del av Zn-variationen föreslås, men inga självklara Zn-former kunde identifieras.

Att tolka matematiska modeller såsom PCA och PLSR och översätta resultaten till förekommande kemiska former kan vara utmanande. Både för PCA och PLSR gäller detta kanske främst uttolkningen av negativa bidrag som ofta föreslås av modellerna, till kemisk sammansättning. Vi föreslår därför att som nästa steg utvärdera statistiska analysmetoder där endast icke-negativa bidrag tillåts, som exempelvis icke-negativ matrisfaktorisering (NMF) eller positiv matrisfaktorisering (PMF). Dessa metoder har inte utvärderats på data inom nuvarande projekt. Utöver att undersöka hur användbara dessa modeller är för ask-data, bör det även tas i beaktning om dessa kan inkludera spatial information på ett meningsfullt sätt.

Baserat på det analyser som gjordes för att utvärdera olika statistiska modellers lämplighet kunde tre huvudtyper av partiklar identifieras:

1. Partiklar med mycket komplex sammansättning, innehållande i stort sett alla analyserade grundämnen. Dessa bedöms vara smältor av partiklar som dras med från förbränningsbädden.
2. Saltpartiklar som sannolikt bildats genom kondensation från gasfas (till skillnad från partiklar som dras med från förbränningsbädden).
3. Kalciumdominerade partiklar, som enligt klusteranalysen förekommer relativt separerade från smältfraktionerna (1).

Zink förekommer i anslutning till alla dessa typer av partiklar, där den sistnämnda gruppen hade lägst nivåer av Zn. Ovan resultat bygger på analyser som primärt användes för att utvärdera metoderna och utvärderingen behöver förfinas och kompletteras innan säkra vetenskapliga slutsatser kan dras.

List of content

1 Introduction

Fly ash from waste-to-energy (WtE) incinerators is typically enriched in metals. While these metals often pose potential threats to the environment and living beings, they also represent a possibility for material extraction.

It is not solely the metal content that is of importance for metal recovery or toxicity, but also the chemical forms of metals as this determines their mobility. An increased rate of metal recovery, achievable from knowing chemical form(s), could both reduce our dependence on virgin materials from mining operation, but also minimize leaching of potentially harmful species to the ecosystem.

Fly ash from WtE is highly heterogenous and contains a mixture of numerous elements of different chemical forms, including amorphous phases. This makes the determination of the chemical form of metals in fly ash challenging. Many metals of interest, such as Cu, Zn, and Pb, are also found in relatively low concentrations (on the order of a few weight percent or lower), further complicating the use of standard methods such as X-ray diffraction.

We have previously shown the potential in using synchrotron based spectroscopical methods in studying chemical form of metals in ash (Gorjatšova et al., 2026; Karlfeldt Fedje et al., 2025; Rissler et al., 2024; Rissler et al., 2020). A key technique to determine speciation in complex matrices is X-ray absorption spectroscopy (XAS), including X-ray absorption near edge structure (XANES) and extended X-ray absorption fine structure (EXAFS). While XANES has proven to be of great use, it requires inclusion of material-relevant reference compounds for proper data evaluation.

As a complementary technique to XANES, seeking to suggest and validate reference compounds included in the analysis, nXRF has been identified as a suitable method. With nXRF the elemental distribution in individual fly ash particles can be analyzed, often presented in the form of 'elemental maps'. While nXRF produces data which at first glance is similar to SEM-EDS data, it is superior in terms of spatial resolution and detection limits. A high spatial resolution and low detection limits are important for metals in the ash particles, as these typically contain nano-sized features.

By analyzing spatial colocalization of elements in nXRF data, we hope to retrieve information on possible chemical forms of metals, and study whether a certain metal is diffusely distributed, or exists at distinct locations in the sample matrix or behaves similarly to other elements in the ash. However, the amount of data generated in nXRF elemental mapping is rather large, presenting challenges in extracting representative information from the samples and utilizing the full potential in the dataset. As an example, a nXRF data set on a 10 x 10 µm sample area may generate over 20 000 spectra (assuming a resolution of 70 nm).

An extensive nXRF data set on fly ash samples from Scandinavian WtE facilities has been collected in a previous project (Karlfeldt Fedje et al., 2025; Rissler et al., 2024). The ash samples studied originates from grate-fired boilers (GB) and one fluidized bed boiler (FB). Within this project the focus is on evaluating ash from the most common boiler type of the two: the GB. The nXRF data was collected at two different beamlines and synchrotron facilities (I13, Diamond Light Source and NanoMAX, MAX IV Laboratory (Björling et al., 2020; Carbone et al., 2022; Johansson et al., 2021)). Initially, the collected data was subject to first-order analysis, i.e. evaluating the elemental maps visually and using simple linear regression analysis. We then found that applying more advanced statistical methods would result in more objective evaluation of the observations and make use of the whole data set in a more efficient way.

Within this project, the aim is to assess various statistical methods for more extensive analysis of the nXRF maps, and to evaluate the selected methods by applying them to the already collected dataset described above. Our analysis will mainly focus on Zn in untreated fly ash, but the aim is that the same approach could be used to study other elements also in treated fly ash samples.

2 Method and implementation

Before evaluating any of the methods, the nXRF data was calibrated, fitted according to standard procedure, formatted, and pre-treated, for example via background subtraction. Elements showing high noise levels in datasets from certain facilities were identified, as these could potentially lead us to charting statistical methods on incorrect grounds. The data was then explored using general visualization tools.

The first order analysis showed that data evaluation requires more advanced statistical methods than a linear regression approach, as data is not only noisy but often also colinear (rank deficient). We thereafter identified several methods of interest based on theory and application, outlined below. These were evaluated on several datasets, with examples presented within this report, focusing on fresh (untreated) fly ash from GB. Our aim is to identify a method that later on can be applied also to treated ashes.

In this context, it is important to note that nXRF data does not give chemical speciation, but as the spatial resolution is high it should be possible to suggest possible compounds based on spatial correlation between elements.

2.1 NXRF MAPPING

Although the name, nXRF, indicates resolution at the nanometer scale, the technique typically offers elemental mapping in 2D with a resolution of 50-100 nm, in the case of our measurements around 70 nm x 70 nm pixel size. The resolution depends on the specific beamline and experimental setup.

By stepwise movement of the sample in relation to the focused X-ray beam, a raster scan of the region of interest (ROI) in the sample can be made. In our dataset each ROI is approximately 10-20 µm, square or rectangular in shape. Using reference samples with known composition and elemental content, it is possible to calibrate the collected spectra to elemental concentrations, and combining this with the spatial information, the resulting data resembles 'elemental maps', as outlined in Figure 1.

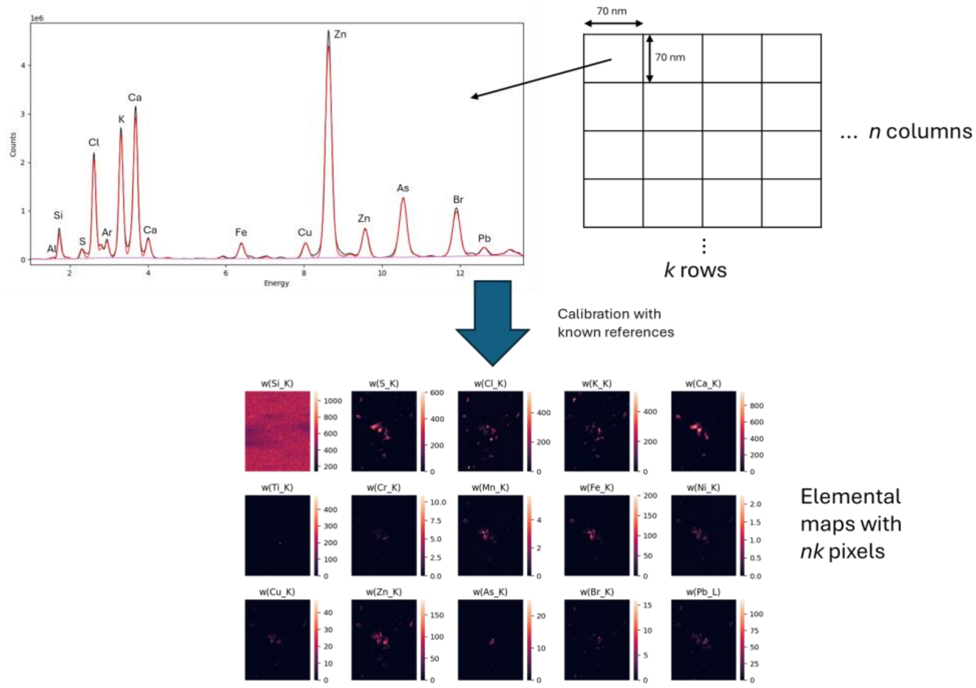


Figure 1. Overview of data handling process producing 'elemental maps'. In each pixel (70 x 70 nm) area in the *n*k* (n number of scanned x-positions and k y-positions) sized raster scanned ROI, an XRF spectra is recorded. Through peak fitting and calibration with known reference samples, this translates to a set of elemental concentrations, *e* in number, in each pixel and a total data set of dimension *n*k*e*.

nXRF synchrotron beamlines typically operate in the range from about five up to ~20-30 keV, depending on the experimental setup. High energy resolution (~2 eV) allows finely resolved spectral peaks with a low detection limit. If using excessively high excitation energy, there is a risk of overlapping peaks around the characteristic energies of interest in the fluorescence spectrum, and the cross section for lighter elements decreases. On the other hand, the incoming energy needs to be sufficiently high for excitation of the elements of interest, in our case targeting Zn (K_α 8637 eV), but also aiming to include Pb. Therefore, we used excitation energies of 13.5-14 keV for the measurements.

At the beamlines where measurements took place, elements heavier than Na can usually be detected. As the fluorescence yield increases with atomic number (and Auger electron emission correspondingly decreases), heavier elements produce stronger fluorescence signals. In contrast, for the lighter elements Auger electron emission is more probable, leading to weaker fluorescent signals. Furthermore, light elements emit lower energy photons, which have a higher probability of being absorbed or scattered before reaching the detector (by the sample itself and by the air between the sample and the detector). As a result, the detected fluorescence from lighter elements is greatly reduced, leading to data with higher noise levels.

In practice, the setup allowed detection of Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Cu, Zn, As, Br and Pb, and in some cases also Se, Bi and Ni. Note however, that due to the reasons given above, the noise levels for the lightest elements (i.e. Al and P) are particularly high. Silicon was not analyzed as thin, X-ray transparent, membranes of SiN were used as sample substrates.

2.2 OVERVIEW AND BRIEF DESCRIPTIONS OF STATISTICAL METHODS EVALUATED

Different explorative and statistical multivariate methods were evaluated for their usefulness in analyzing spatially resolved elemental data. The overall aim is to present a suitable method that can suggest possible chemical composition of fly ash samples, based on colocation of elements, specifically targeting Zn colocation.

2.2.1 Back-mapping trends seen in scatterplots

In the first-order analysis, with linear regression, we noticed that in some element scatterplots there were several distinct trends visible (within one ROI), as seen in the example in Figure 2. While the correlation plots reveal these distinct trends, Pearson’s correlation coefficients are unable to reflect complexities like dual or mixed trends (e.g. Zn vs Cr).

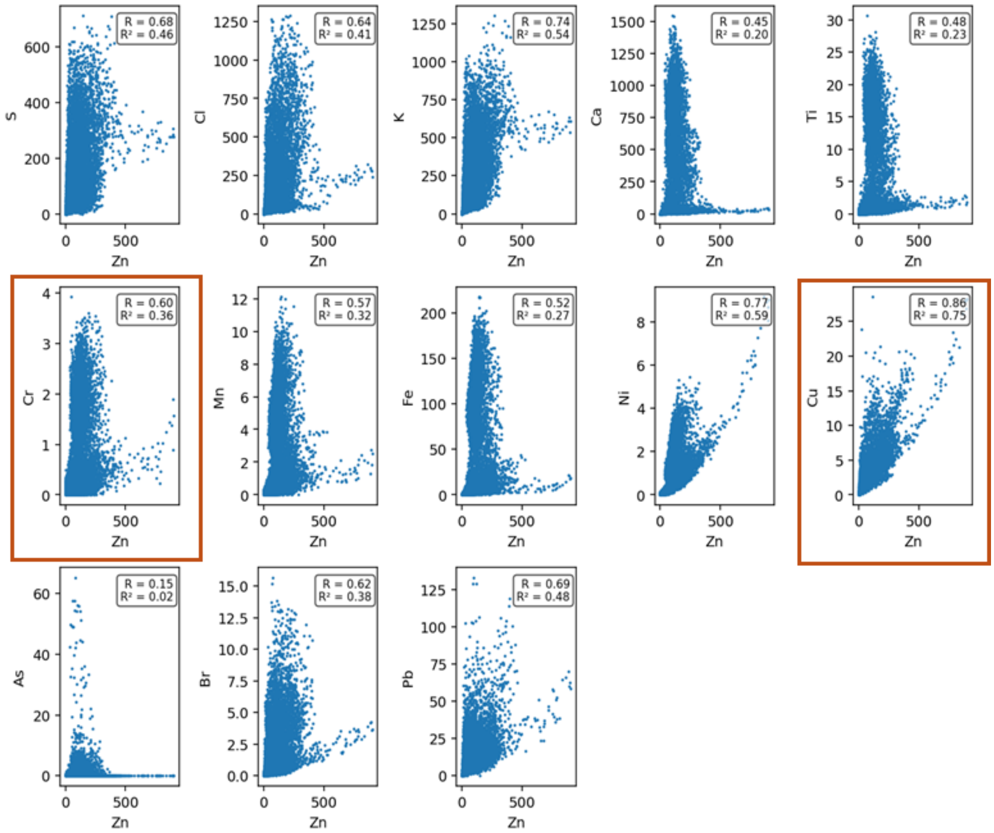


Figure 2. Example of scatterplot from nXRF data where two distinct trends are visible in for example the Zn vs Cr and Zn vs Cu subplot. Each data point corresponds to one pixel in one ROI from a grate-fired boiler sample.

In order to explore these patterns, one way is to select pixels belonging to one of the trends at a time, and back-map these onto the original elemental maps, similarly to what was done in a previous publication on soil samples (Muehe et al., 2013). This is not to be viewed as a strict statistical approach but may be used to visualize trends in specific ROIs that can be further explored. This approach can suggest some interesting relations and improve understanding of co-located elements in the ash particles, although on more case-study basis compared to more advanced statistical approaches. The representativeness of such a ROI and the selection might be subjective.

2.2.2 Cluster analysis (HDBSCAN)

Cluster analysis can be used to group datapoints based on similarity in the selected dimensions. In our case, this translates to similarity in elemental composition between different pixels. Which, and how many, different compounds that are present in each ash sample are unknown. Therefore, we choose an unsupervised clustering method called hierarchical density-based spatial clustering of applications with noise (HDBSCAN). A simpler clustering method, namely DBSCAN, was first evaluated. However, as this method was sensitive to parameter selection and did not yield meaningful results, we chose to proceed with HDBSCAN instead. HDBSCAN is a useful method for explorative data analysis since it requires minimal parameter tuning, does not assume a certain cluster distribution (it's nonparametric), and can handle noise data points separately from clusters (McInnes & Healy, 2017). HDBSCAN finds dense regions without a set density threshold, but rather creates a clustering hierarchy from an infinite range of density thresholds (Campello et al., 2015).

Prior to clustering, the background, here meaning areas in the ROI without Zn-containing particles, was removed below a set threshold value for Zn concentration. After this, each ROI was normalized pixelwise, enforcing the sum of all elements present to equal one. Background filtering ensures that only Zn-containing particles are analyzed, and normalization makes sure that any effects due to particle thickness are diminished.

The major analyzed elements (S, Cl, K, Ca, Fe, and Zn) were used as clustering variables. All variables were standardized (centered around zero and scaled to unit variance) before clustering, to allow analysis of elements on different concentration scales.

The HDBSCAN algorithm seeks clusters by identifying groupings in data, in how densely the data points are distributed in the feature space, spanned by the selected dimensions, and treating each pixel as an observation. The algorithm utilizes a threshold for allowed within-cluster distance and builds a hierarchical cluster structure by progressively varying the density threshold. In the final step the algorithm investigates the stability that the clusters have over thresholds, such that only clusters that remain coherent over a wide range of density thresholds are considered meaningful and are selected in the final output. This approach enables HDSCAN to naturally identify clusters of various shapes and sizes, while also detecting noise points as those that never form part of a persistent cluster (referred to as cluster -1).

Energiforsk

In more detail, the algorithm starts with a local measure of density, called the core density = the distance to its k^{th} nearest neighbor (which is small in dense areas and large in sparse areas). The core distances are incorporated into a 'mutual reliability distance' by thresholding the distance between the points by the core densities. This raises the distances where necessary, to reflect the difficulty of connecting two points through high-density regions.

To determine the stability of clusters, the parameter λ is introduced, where λ =1/mutual reachability distance. A minimum spanning tree is constructed from the mutual reachability distance by successively linking points, avoiding already connected points. The minimum spanning tree is then condensed (removing very small or short-lived clusters), before final clusters are selected based on their stability, i.e. how long they persist across different λ values.

The main parameter to set in HDBSCAN is minimum cluster size, simply corresponding to how many pixels are needed to form a cluster. Additionally, the parameter minimum samples may be adjusted (by default minimum samples = minimum cluster size), controlling the number of neighbors to a core point in the algorithm. Assigning minimum samples to a lower value than minimum cluster size will in practice lead to fewer data points being seen as noise, as the cluster density requirement is not as stringent. These parameters were varied when evaluating the clustering method on the fly ash nXRF data.

The spatial information was not included in the clustering analysis, meaning that clusters may form between pixels that are not in physical vicinity within the sample. The reason for this approach was that it is difficult to know how to scale the effect of spatial distribution compared to similarity in elemental composition between the clusters, which is critical as this could affect how the model handles clustering of differently sized particles.

2.2.3 PCA

Principal component analysis (PCA) is a multivariate latent-variable model commonly used for dimensionality reduction and is often useful when data shows multicollinearity (Jolliffe & Cadima, 2016). As the nXRF data have a rather large number of variables (i.e. elements), reducing the data into a smaller set of orthogonal components (principal components) can be useful to discover elemental patterns of covariation.

Each principal component (PC) is a linear combination of variables (here elements) that is selected to explain as much of the variance in the observations as possible, in our case varying elemental composition in different pixels. The first PC will describe most of the overall variance seen in the data, while PCs with higher numbers represent factors that affect overall variance less. PCA is an unsupervised method, meaning that it cannot be used to specifically target Zn. PCA does not formally require any specific underlying distribution, but it is important that the relationships expected are linear.

Prior to applying PCA, background areas not containing particles (here defined by low Zn content) were removed. Data on different scales is standardized. The original data matrix will be decomposed into loadings, describing how each variable (element) contributes to the PCs, and scores, corresponding to observations in the PC transformed space. A large value in the loading vector of an element (variable) indicates that it contributes strongly to that PC. Variables whose loading vectors point in a similar direction (for example Cl, Br and, K in Figure 7) are positively correlated, while variables pointing in opposite direction (for example As and S in PC3) are negatively correlated in the specified PCs.

2.2.4 PLS

Partial least squares (PLS) is, in contrast to PCA, a supervised model. This means that it seeks latent variables that are valuable for explaining the covariance of a specific response, in our case Zn, with the other elements (the predictors). The most important latent variable will be the one best describing not only the overall variance in the data-set (as in PCA) but the one that also have a high correlation to the response variable. A mathematical description can be found for example here (Rosipal & Krämer, 2006).

PLS is a family of methods, where PLS regression (PLSR) is the most commonly used version for continuous variables. To understand how different elements covary with Zn we evaluate the PLSR weights. PLSR weights describe how each element is combined into the respective latent variables. As an example, if element A and B have high weights for the first latent variable, this could be interpreted as both A and B in combination can explain rather large parts of the Zn concentration variation, which can be interpreted as that A and B possibly form a compound with Zn. One major difficulty in doing such direct translation between the mathematical model and its physical interpretations is that PLSR, as PCA, can produce negative weights.

Because PLSR models both predictors and response jointly, through shared latent variables, the importance of individual predictors cannot be read directly from the regression coefficients. Instead, the common measure variable importance in projection (VIP) score is introduced. The VIP score summarizes how strongly each original X-variable contributes to the latent components, weighted by how much of these components explain the variance of the response, Y (here Zn). Thus, VIP reflects the combined influence of a variable on both the X-structure (via its PLS weights) and the predictive power of the components on Y. Variables with VIP scores greater than 1 are generally considered highly influential. The VIP score is one example of how PLSR result interpretation may require more effort compared to PCA, but in return the PLSR is often more powerful for targeted analysis.

By predicting Y for a subset of observations at a time (we use 10-fold cross-validation), the latent variables' ability to describe the variation in X that is relevant for Y can be evaluated via the cross-validated coefficient of determination (R^2), known as the Q^2.

Only a brief summary of the algorithm and mathematical relations are given here, details can be found in (Wold et al., 2001). PLSR builds a linear multivariate regression model capturing the covariance structure between the predictor matrix X and the response variable Y, to model both X and Y by the same set of latent variables. To construct the PLSR model, an algorithm (NIPALS) sequentially extracts latent variables by finding linear combinations of the predictor (X) and the response (Y) variable(s), described by their weight matrices, which maximizes their covariance. The loading matrices, which are used in the data reconstruction and deflation, are derived via least-squares projection.

3 Results

3.1 BACK-MAPPING TRENDS SEEN IN SCATTERPLOTS

Selecting pixels followed by back-mapping to separate the different trends seen in scatterplots of the data is one way to retrieve information about fly ash particle composition. This is illustrated by plots from an example ROI, shown in Figure 3. In this ROI there are pixels with high variability in Cu, but close to no covariation with Zn, and low overall Zn levels, (red datapoints in left panel), while other pixels show a clear correlation between Cu and Zn (grey in left panel).

Selecting the sub-set of pixels for which no correlation between Cu and Zn were observed (red), these turned out to correspond to a specific particle on the elemental maps (lower left panel in Figure 3). In the Br-Zn scatterplot these selected data points/pixels show a high Br content, for which the Br concentration also varied almost independently from Zn.

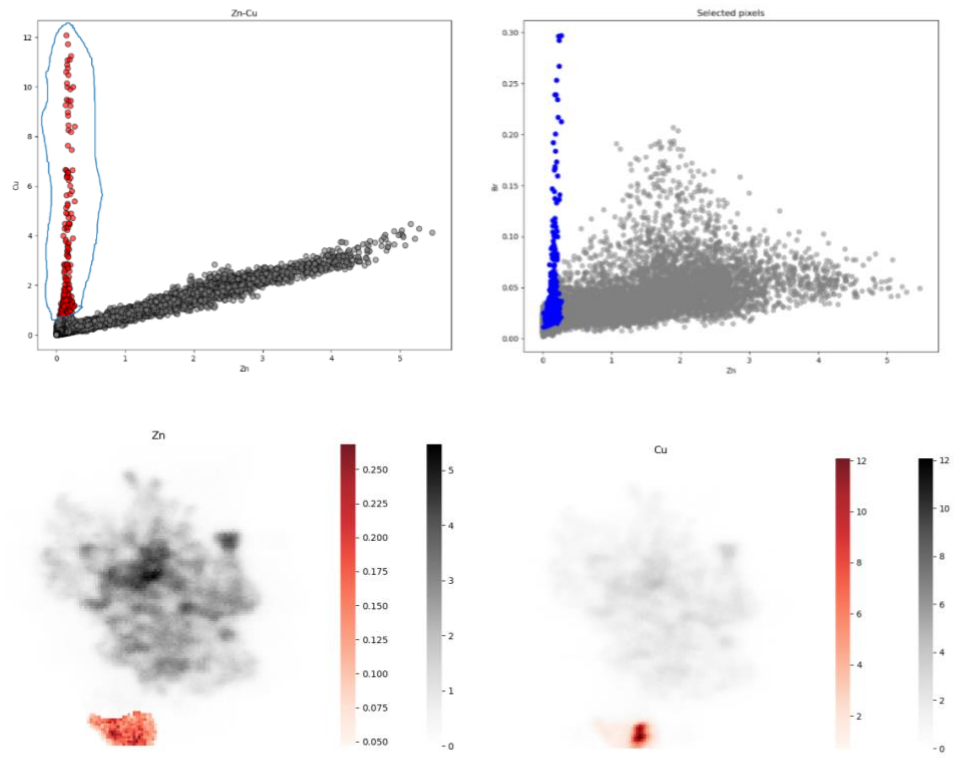


Figure 3. Example of back-mapping selected pixels from scatterplot. Upper-left: Zn-Cu scatterplot where selected points are marked in red. Upper-right: Zn-Br scatterplot, where points marked in blue correspond to the ones selected in the Zn-Cu plot. Lower-left: Zn map with overall ROI concentration in grey-scale and selected points in red (note the differentscale). Lower-right: the corresponding Cu map.

Instead selecting the other set of pixels in the ROI, for which there was a clear correlation between Zn and Cu, some trends become clearer (examples of scatter plots are shown in Figure 4). The major drawback with this type of analysis is that the data treatment is rather manual.

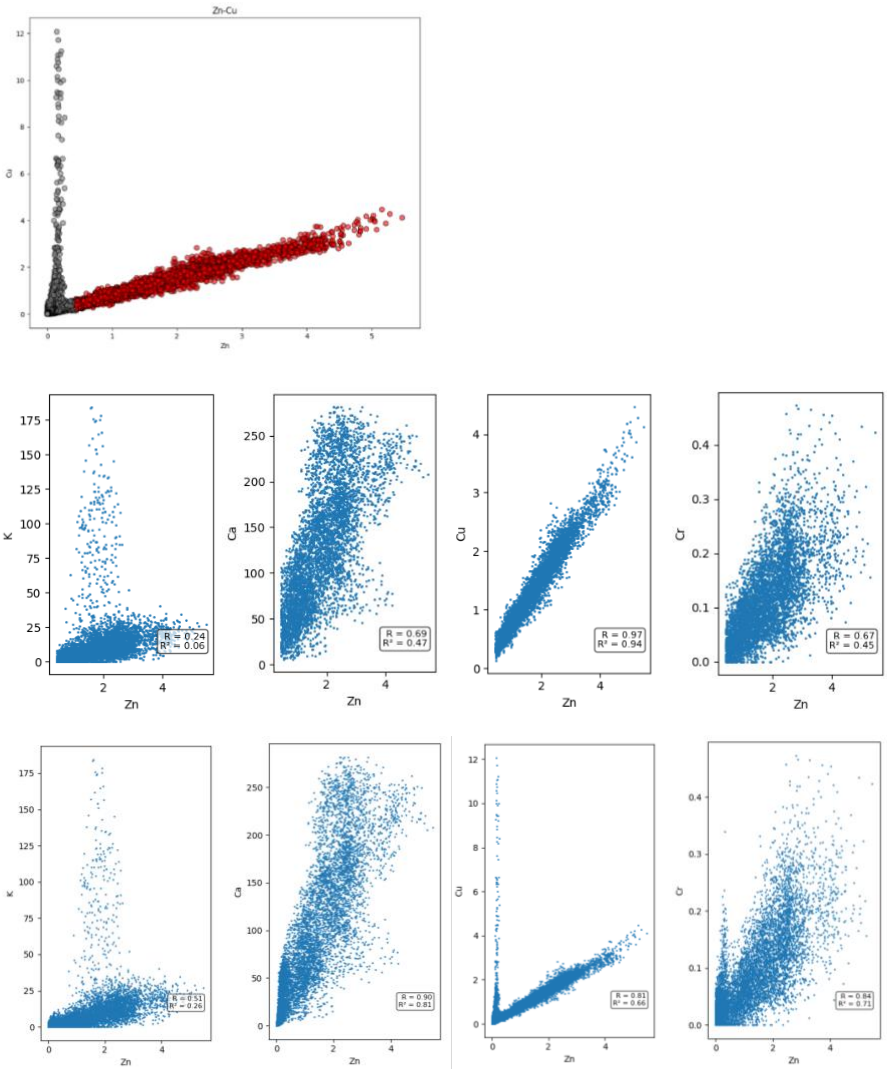


Figure 4. Example of selecting pixels from scatterplot (same ROI as above).
Upper: Pixels selected from Zn-Cu scatterplot marked in red.
Mid: Scatterplots of Zn vs selected elements only including selected pixels.
Lower: Scatterplots of Zn vs selected elements including all ROI.

Energiforsk

3.2 CLUSTERING (HDBSCAN)

HDBSCAN clustering was implemented for several samples but will be illustrated using a grate-fired boiler fly ash. In the analysis all ROIs from the sample were included, resulting in a total of 32 160 pixels after filtering (background subtraction). Clustering was based on the six elements of highest concentration (Ca, Cl, K, S, Fe and Zn) of those measured in the XRF. The number of variables (elements) used for the clustering should be limited, as increasing dimensionality decreases contrast between dense and sparse regions. Aluminum and P were also major elements, but as the noise levels were considered high, these were not included in the analysis. In the example given, a minimum cluster size of 30 pixels was used.

In the analysis results, more than half of the pixels were classified as noise, regardless of parameter tuning, and the noise cluster had a quite high relative Zn concentration. The results are shown in Figure 5. The fraction of Zn in each cluster was calculated from original data, accounting for S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, As, Br, Pb and Zn as the total of 'all' elements, given in figure text (Figure 5).

Comparing the identified clusters, the largest cluster (cluster 6, n=9,924) contains high levels of K, S, Cl, but also Zn, and seems to represent easily soluble salts, likely formed by condensation. The second largest cluster (cluster 4, n=2,432) contains pixels dominated by Ca and S, interpreted as possible $CaSO_4$. A third large cluster identified (cluster 3, n=2109) mainly contains Ca with some minor contribution from all other elements included in the clustering. This cluster could be $CaCO_3$ or CaOH but also compounds with Si and/or Al as those elements are not included in the analysis. For example, gehlenite has been found in fly ash (Gorjatšova et al., 2026) and is further known to form solid solutions with transition metals such as Zn.

The highest relative concentration of Zn is found in clusters 0 and 1. In cluster 1, Zn made up 15% of the mass and the cluster was dominated by Zn, K and Cl, which is in-line with that K_2ZnCl_4 has been reported in fly ash (Gorjatšova et al., 2026). Cluster 0 had 21 wt % Zn and was more mixed with Ca and S, along with K, Cl and Fe. Note however that the number of pixels in this cluster is relatively low.

The concentrations in the nXRF dataset are mass-weighted. Because Zn is the heaviest element included in the cluster analysis, its contribution decreases when the data are converted to number-weighted fractions. Consequently, none of the clusters represents any pure Zn compound. Not so surprisingly, all clusters contain mixtures with other elements and non-Zn phases.

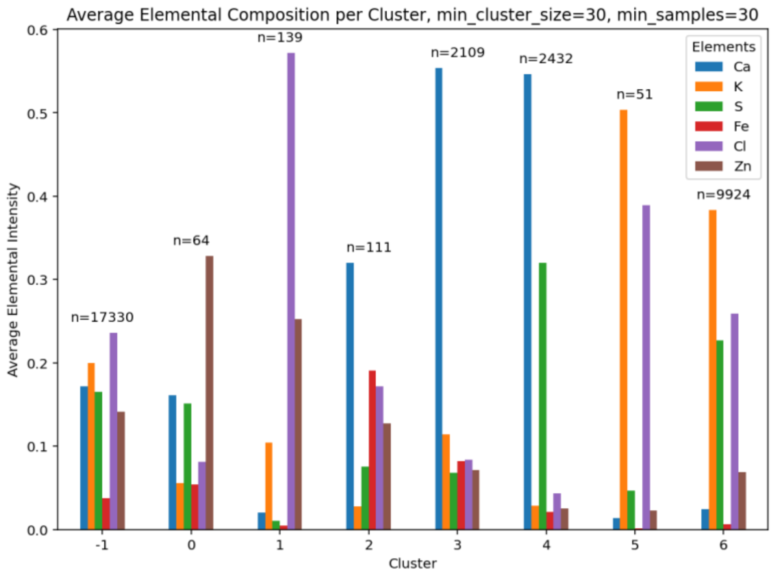


Figure 5. Grate-fired boiler fly ash clustering example. Average normalized concentration of elements used in the clustering, separated by their cluster belonging (HDBSCAN). In the graph, the pixels classified as noise are labelled as -1. The fraction of Zn in respective cluster was for cluster -1 - 10%, cluster 0 - 21%, cluster 1 - 15%, cluster 2 - 11%, cluster 3 - 7 %, cluster 4 - 4%, cluster 5 – 3% and cluster 6 - 7%.

While HDBSCAN clusters pixels based on similarity in elemental composition, it does not directly give information on correlations between elements. Clustering can rather be viewed as a method for grouping data, which can then be subject to further analysis. To elaborate on clustering efficiency, we looked at scatterplots of Zn vs the other elements within the different clusters. As an example, we present the largest cluster (Cluster 6) identified in the example above, in Figure 6. While some trends have become clearer, data within the cluster still shows a high variation, indicating that this cluster does not correspond to Zn in a single chemical form.

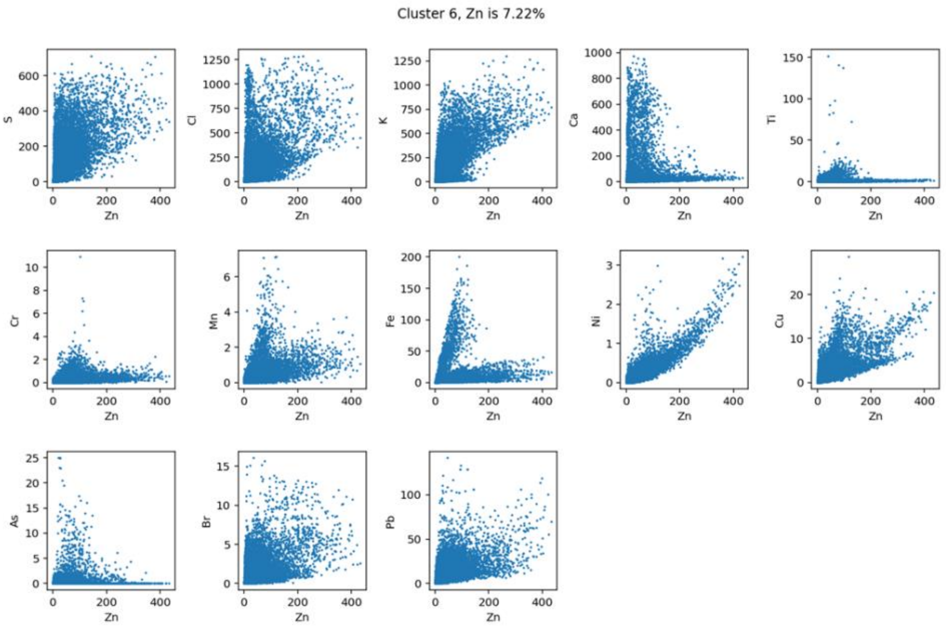


Figure 6. Example of Zn scatterplots in a single cluster, using the largest cluster identified by HDBSCAN.

The clustering approach seems to be helpful to some extent, as it can group pixels to find clusters with similar elemental composition, which may represent chemical forms present in fly ash. However, the majority of the pixels were classified as noise and could not be sorted into any cluster. As the HDBSCAN clustering is based on similarities in the elemental composition in different pixels (i.e. density), we believe that it struggles to find clusters in a data set with many pixels that represent mixtures of different compounds fused together and in varying fractions over the pixels. As an example, if compound A and B each represent 50% in some pixels, and in others their proportions are 10/90%, these two collections of points would not be identified as belonging to the same cluster. The same goes for example if smaller "salt dominated" particles are agglomerated on top of a larger particles of different composition.

In summary, clustering approaches work best if different compounds are not mixed in the same pixel. A conclusion from our evaluation of cluster analysis for fly ash particles is that, in most cases, this does not seem to be the case; rather, the fly ash particles are composed of mixtures of different compounds in varying ratios. However, when the model succeeds in identifying clusters, it may provide useful information by showing which elements are associated with Zn within those clusters and by indicating that particles with this elemental composition often occur in less complex mixes. It should be kept in mind, though, that with this approach, the major Zn species may be overlooked if they typically occur in complex, varying mixtures.

3.3 PCA

PCA is entirely unsupervised, meaning that it is not a method that can be targeted to explain a specific element, such as Zn. The first principal components will be dominated by the variable combinations describing the highest variance in the data set. As this is not necessarily Zn, a PCA may be helpful in explaining the main compounds formed in the ash, while suggesting Zn chemical form from the resulting components may be challenging.

To illustrate the PCA analysis of the nXRF data, we use a dataset (eight ROIs) of ESP ash from a grate-fired boiler, covering both coarse and finer particles. This selected dataset includes the elements S, Cl, K, Ca, Ti, Cr, Mn, Fe, Cu, Zn, As, Br, and Pb. The PCA is, in this case, based on four principal components.

The first principal component had the highest loadings for Mn, S and Zn, followed by Cr, Ca, Cu, and Fe. All loadings for the first principal component were positive, which is likely an effect of the fact that this component explains the variance in concentration (non-negative) over the pixels and reflects particle thickness rather than chemical associations.

The second and third principal components, presented in Figure 7, describe variance that was not accounted for in the first component, and could possibly reveal more subtle ash particle compositional information. For example, K, Br, and Cl, all display large loadings in PC2 (seen as long vectors) and pointing in a similar direction (Figure 7), consistent with a shared chemical association, possibly related to condensation of different salts. Zinc, on the other hand, has a short loading vector for PC2 and PC3, indicating that relatively little of the variance not captured by PC1 seems to be describable by Zn. Overall, this suggests that PCA might be useful to understand ash composition in general but may not be the best option for our purpose, as it does not target Zn.

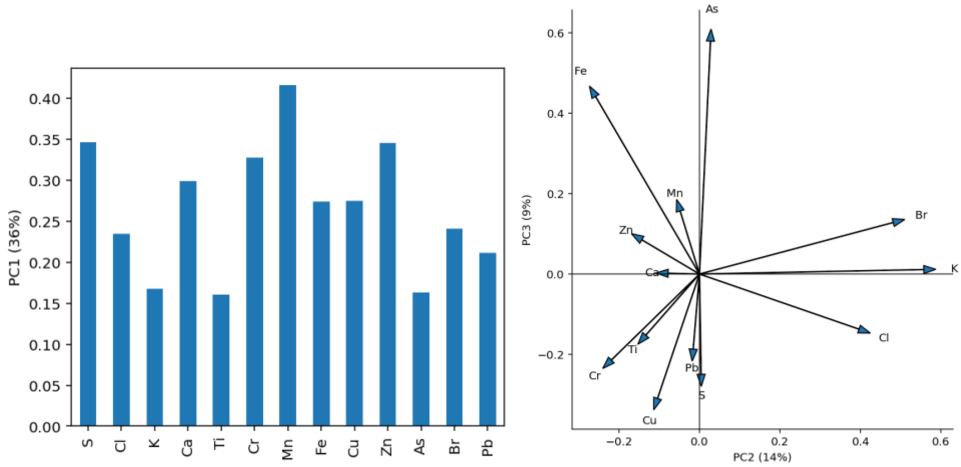


Figure 7. Loadings in PCA, the number in parenthesis is the amount of explained variance for that component. Left: Loadings for the first principal component. Right: Loading plot of the second vs. third principal components.

3.4 PLSR

PLSR results will be illustrated using the same dataset as for PCA. First, pixels not containing Zn were removed (background removal). Thereafter a standardization of the X variables (S, Cl, K, Ca, Ti, Cr, Mn, Fe, Cu, As, Br, Pb) was done. From Q^2 we concluded that two latent variables were sufficient to capture most information, with an increase in Q^2 by less than 0.02 if a third latent variable was included, as shown in Figure 8 (i.e. it did not help much in terms of explaining Zn variation).

VIP scores (explained in Section 2.2.4) revealed that that Mn and Fe were particularly important to predict Zn concentration in the ash particles, but also that S, Ca, Cr, Cu and Pb were important, Figure 8.

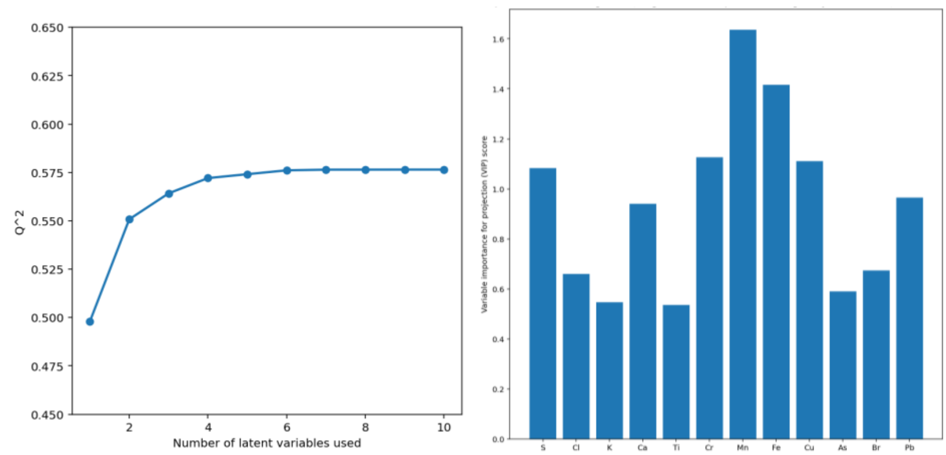


Figure 8. PLSR example. Left: Q^2 as a function of number of latent variables included in the model (notice y-axis scaling). Right: VIP scores for the different elements for a PLSR model with two latent variables.

More detailed information on how different elements covary with Zn according to the PLSR model can be found through studying element weights in each latent variable identified by the model, shown in Figure 9. For the dataset used here, the first latent variable was made up solely by positive contributions from all elements analyzed, of which Mn and Fe had slightly higher weights. This can also be seen when comparing the score map of the first latent variable with the original elemental maps in one example ROI included in the dataset, shown in Figure 10. Areas with high intensity in the first latent variable often correspond to areas with high concentration of either Zn, Mn or Fe, despite the analysis taking many ROIs into account when finding the weights that yield maximum covariance. As for PC1 in the PCA, it is possible that this first latent variable captures the effect of sample thickness rather than composition. Covariation with Zn is likely stronger in areas where ash particles are thicker, which could explain why the first latent variable had positive weights for all elements. However, it could also indicate that Zn is present in forms associated with close to all elements analyzed and thus will covary to some degree with all elements.

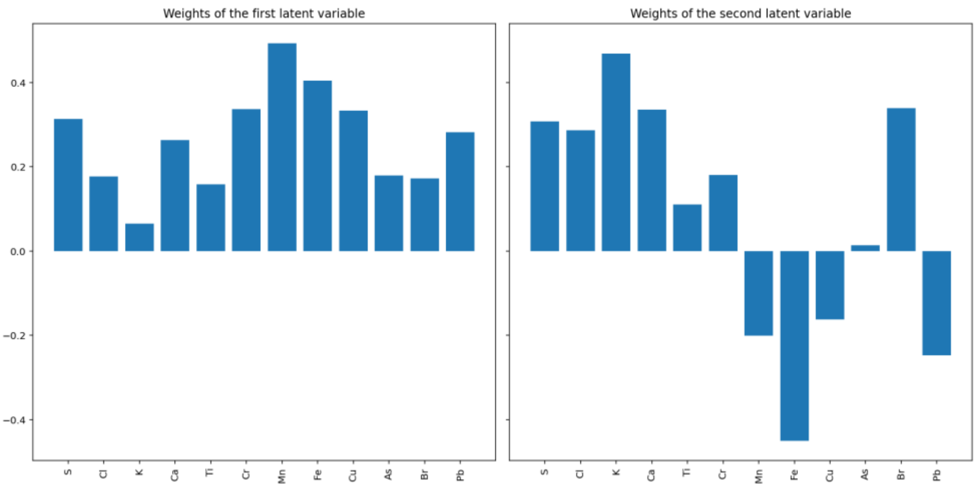


Figure 9. PLSR model with two latent variables. The weight of each element is shown separately for the two latent variables.

The second latent variable had a strong positive weight for K together with S, Cl, Ca and Br, while Fe, Pb, Mn and Cu had negative weights. From comparison of the scores and the elemental maps (Figure 10), the second latent variable seems to correspond to more peripheral areas in the ROI with smaller ash particles/agglomerates. This is in-line with the expectation that the condensates, typically composed of Na, Cl, K, and Br, form smaller particles that are expected to adhere to the surfaces of larger particles.

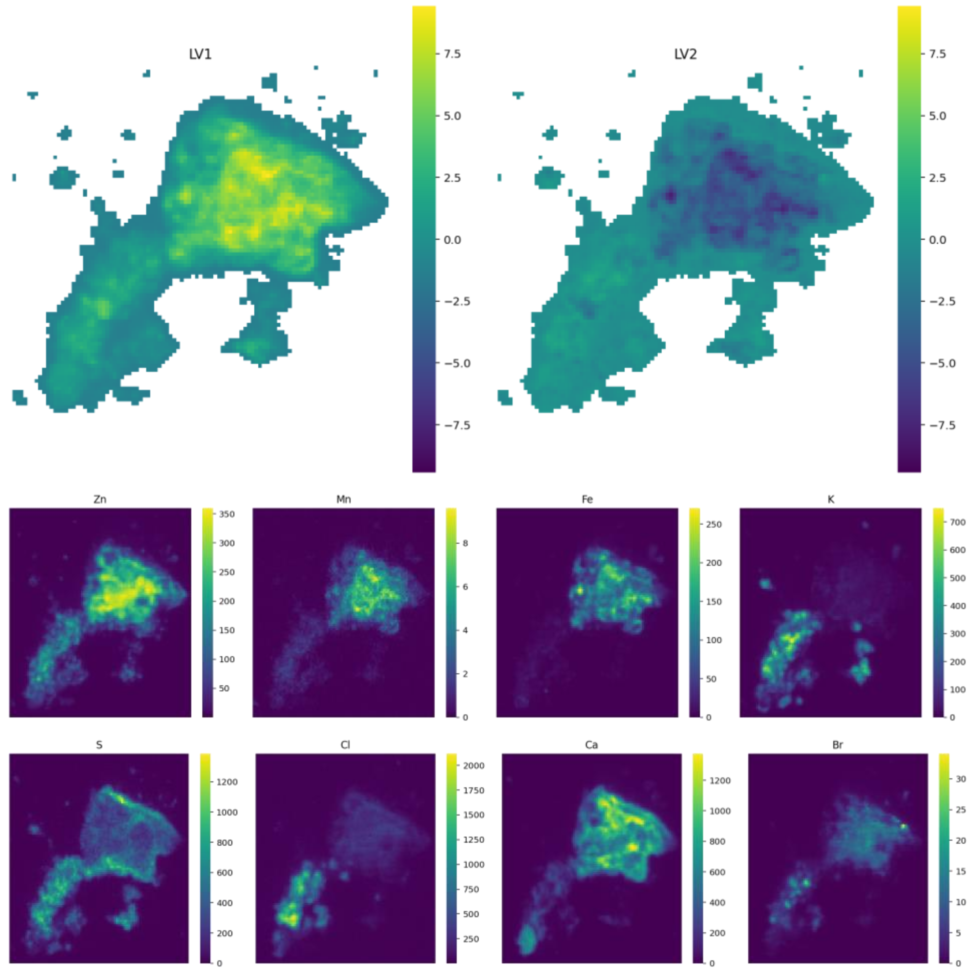


Figure 10. Elemental maps. Upper: PLSR scores for first and second latent variable, describing how each pixel is represented. Lower: Original nXRF elemental maps for Zn, Mn, Fe, K, S, Cl, Ca and Br.

A few elements (Se, Bi, Al, P) were excluded from the PLSR analysis presented above, as they were not available in all data sets. When we include these in the analysis, trace elements such as Se improved the model predictive capabilities, increasing Q^2 from 57% to >68% when four latent variables were included. Four latent variables were included in this case as Q^2 was clearly improved by including more than two latent variables. In relation to this, it is important to stress that high predictive power for Zn does not necessarily imply chemical binding between elements, especially for elements only occurring in trace amounts such as Se. In the case of Se, we think it may reflect a similar behavior of Se and Zn during particle formation.

As the PLSR analysis on multiple ROIs (of ash from the same sample) did not reveal any distinct element(s) that consistently covaried with Zn, the PLSR model was instead fitted to single ROIs to see if the model could pick up trends that we found visible directly from the elemental maps. As an example, we here showcase the results from a ROI, where the distribution of Zn, Cu, K and Fe is shown in Figure 11. This specific ROI shows a spherical metal particle with clear signal from Fe, Mn, Cr, Ti and Ca, in the figure represented by the Fe map. The metal sphere is surrounded by aggregated smaller particles composed by elements that are known to exist at least partially in volatile forms, Zn, Cl, K, S, Cu, Br and Pb (a selection of elements shown in the figure).

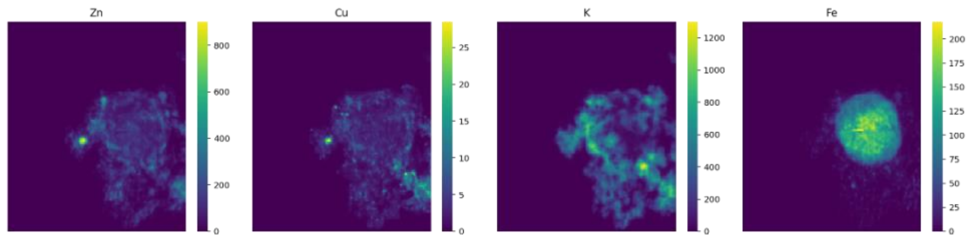


Figure 11. Example of single ROI PLSR analysis, elemental maps of Zn, Cu, K and Fe, showcasing a spherical metal particle with fine particles aggregated to the surface (condensates). Cl is distributed similarly to K.

The highest PLSR (here including four latent variables in the model) VIP by far (not shown) in the single ROI case (Figure 11) was Cu, followed by Pb, K and S. Again, the first latent variable, shown in Figure 12, consisted of only positive weights with relatively similar contributions from many elements (similar to the multiple ROI example). The second latent variable had a strong positive weight for Cu and smaller, yet positive, weights for Pb, with scores corresponding well to Cu and Zn spatial distribution in original maps. The third latent variable had positive weights for elements that are expected to be found in the smaller particles formed by evaporation-condensation in the flue gas, such as S, Cl, K, Br and Pb. While this example illustrates that PLSR can in certain cases pick up some trends (localized Cu spots and particle areas with more volatile elements) from high resolution spatial data, it seems to be of limited use for understanding larger data sets with more varying mixes of Zn species.

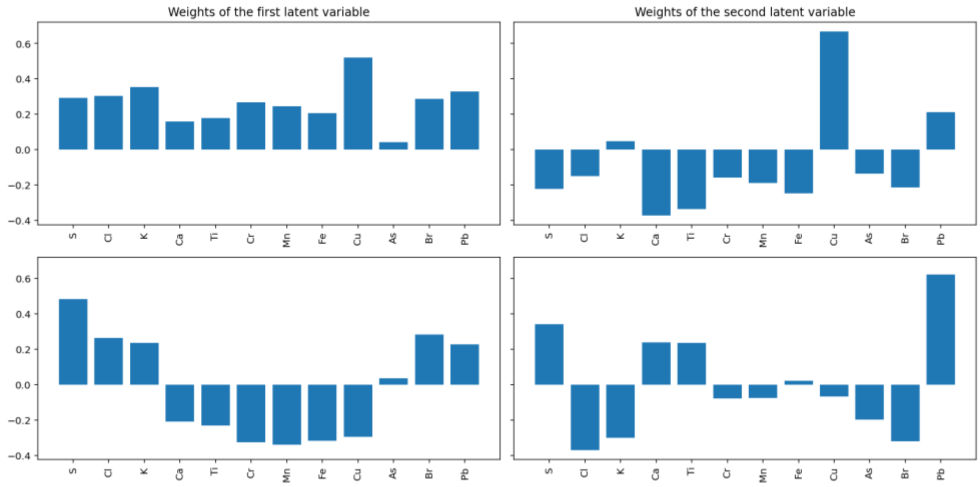


Figure 12. Weights of the latent variables in the single ROI example.

Although PLSR is a great tool in terms of handling collinearity, the results are in many cases difficult to interpret physically. While the addition of latent variables with different weights makes sense mathematically, it was hard to distinguish clear elemental patterns based on the resulting latent variables. This could partially be a consequence of how the model calculates later latent variables based on covariation that is not captured by previous latent variables, sometimes leading to negative weights, which are hard to interpret physically. Another challenge is that Zn possibly is present as solid solutions, with potentially varying ratios of Zn to other elements.

4 Model choice and relevant methods not yet evaluated – NNMF and PMF

Finding an optimal analysis approach for spatially resolved elemental data of ash particles is challenging, especially if the element of interest is low in concentration, as is the case of many transition metals. Due to the presence of fused mixed phases, often in varying ratios, statistical methods which can combine features (ultimately corresponding to chemical phases) in a weighted manner may be preferrable. Having each pixel be described by a combination of different features can be referred to as a distributed representation and is common in factor analysis group of methods, e.g. PCA, PLSR. Contrasting to this, cluster analysis normally does not allow a distributed representation, since each pixel can only be assigned to one cluster, which may not allow the full complexity in varying-ratio-mixed ash particles to be represented.

The distributed representations methods evaluated within this project, PCA and PLSR, both have the disadvantage that they can result in negative feature weights, which, as previously mentioned, is not readily interpretable in terms of chemical compositions. An alternative method which allows distributed representation while avoiding negative weights is non-negative matrix factorization (NNMF) (Lee & Seung, 2000; Lee & Seung, 1999). Although commonly used for other applications, the potential usefulness of NNMF has, to the best of our knowledge, not yet been shown for this type of spatial multi-element data, despite the existence of more advanced extensions of NNMF which can take also spatial information into account in identifying factors. The spatial aspect is in fact not addressed in any of the methods evaluated within this work (see discussion for cluster analysis).

A related method, freely available via an interactive GUI through US EPA, is positive matrix factorization (PMF). PMF is common in environmental analysis, mainly due to its interpretability. Compared to NNMF, PMF requires an uncertainty matrix input, which needs to be carefully considered as it is highly influencing the output (Paatero & Tapper, 1994). In contrast to standard NNMF, PMF supplies tools that can help reduce rotational ambiguity, which increases the chance of translating model output to interpretable chemical composition. However, the ordinary use-case for PMF is temporally rather than spatially resolved data. Similar to PCA, NNMF and PMF is unsupervised in the sense that the models explains the variability in the dataset in general and is not directed towards explaining any specific element.

5 Conclusions

The aim of this project was to identify and evaluate data analysis methods suitable for elemental microscopy data, such as that from nXRF, typically containing a large number of spatial positions (pixels) with elemental concentrations.

Manual selection of pixels, based on visual trends in scatterplots, was used to study correlations between different elements in the selected pixels and to visualize the position of selected pixels in their original elemental maps. The manual selection had the advantage of avoiding mathematical transformation and allowed straightforward interpretation. This method may provide insight into composition of individual particles/clusters of pixels, but is time-consuming and may be biased, particularly as ROIs with clear trends (in comparison to more complex patterns) in the elemental scatterplots may be favored for this type of hands-on analysis.

Cluster analysis groups pixels based on their similarity in elemental composition, here using the main elements measured. Clustering, more specifically HDBSCAN algorithm, resulted in a few clusters with physically interpretable elemental compositions for the datasets evaluated. The main issue with this method was that more than half of the pixels were categorized as noise due to complex and varying elemental composition in the pixels, leaving only half of the data points interpretable. The cluster assignment is problematic analyzing pixels containing particles of different elemental composition, when present in varying ratios in the dataset/pixels. This could for example be aggregated particles, components fused together in the high temperature process into a particle, or metals substituted into solid solutions to varying degrees in different particles. Still, this type of method can provide important insights for understanding the chemical forms of metals in the ash but will not lead to any unambiguous suggestions of Zn species, since the model may only assign each pixel to one specific cluster, alternatively treat it as noise.

Due to the fused mixed phases often present in the ashes, statistical methods that can combine features (ultimately corresponding to chemical phases) in a weighted manner are promising. Having each pixel be described by a combination of different features is a common trait for factor analysis methods, e.g. PCA and PLSR.

PCA was evaluated and shown to be useful in providing the major compounds present in the ash through analysis of the elemental loadings in each component, which in the best-case scenario represents distinct chemical species. However, PCA cannot be forced to directly target an element of interest and helped little in explaining Zn colocalization patterns in the particles.

To specifically target Zn in the analysis, we evaluated another factorization method, PLSR. Analyzing the PLSR weights, it was possible to suggest trends that might explain some of the Zn variation, but no obvious suggestions for Zn chemical forms could be made.

The main challenge with PCA and PLSR was how to interpret negative weights in terms of elemental contributions. Therefore, we suggest that future efforts should be made addressing the usefulness of NNMF and/or PMF. Besides the fact that these models do not introduce negative elemental weights, NNMF can be made incorporating spatial information, in which we could see potential if implemented in a careful way. These models are, however, not targeted at describing any specific element, and will likely need to be combined with a second step in the analysis to target colocalization with Zn specifically.

Table 1. Overview of methods evaluated and their applicability to the test data.

Method	**Pros**	**Cons**
Manual pixel selection and back-mapping	Illustrative Straight forward interpretation	May be biased towards selection of clearly visible trends. Complex patterns not captured. Full data set not analyzed
Cluster analysis (HDBSCAN)	Interpretable information about chemical composition of clusters	Most pixels not classified (complex mixtures)
PCA	Shows trends in major composition of fly ash particles	Not addressing Zn specifically Negative loadings challenging to interpret physically
PLSR	Directed towards specific element (Zn)	No obvious trends predicted by the model Negative weights, resulting from complex mathematical model, challenging to interpret physically

6 References

Björling, A., Carbone, D., Parfeniukas, K., Kahnt, M., Kalbfleisch, S., Sala, S., Johansson, U., & Vogt, U. (2020). Ptychographic characterization of a coherent nanofocused X-ray beam. *Optics Express, Vol. 28, Issue 4, pp. 5069-5076*, *28*, 5069–5076 , pmid = 32121735 , publisher = Optica Publishing Group. https://doi.org/10.1364/OE.386068

Campello, R. J. G. B., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Trans. Knowl. Discov. Data*, *10*(1), Article 5. https://doi.org/10.1145/2733381

Carbone, D., Kalbfleisch, S., Johansson, U., Björling, A., Kahnt, M., Sala, S., Stankevic, T., Rodriguez-Fernandez, A., Bring, B., Zdenek, M., Bell, P., Erb, D., Hardion, V., Weninger, C., Al-Sallami, H., Lidon-Simon, J., Carlson, S., Jerrebo, A., Jensen, B. N.,…Roslun, L. (2022). Design and performance of a dedicated coherent X-ray scanning diffraction instrument at beamline NanoMAX of MAX IV. *urn:issn:1600-5775*, *29*, 876–887 , pmid = 35511021 , publisher = International Union of Crystallography. https://doi.org/10.1107/S1600577522001333

Gorjatšova, E., Bergman, F., Clauss, K. G. V. S., Skoglund, N., Fedje, K. K., & Rissler, J. (2026). Zinc speciation in fine and coarse fly ash particles collected in-flight at a waste incinerator. *Environmental Au*. https://doi.org/10.1021/acsenvironau.5c00222

Johansson, U., Carbone, D., Kalbfleisch, S., Björling, A., Kahnt, M., Sala, S., Stankevic, T., Marianne, L., Angel Rodriguez, F., Björn, B., David, P., Karina, T., Paul, B., David, E., Clemens, W., Zdenek, M., Linus, R., Karl, A., Brian Norsk, J.,…Ulrich, V. (2021). NanoMAX: the hard X-ray nanoprobe beamline at the MAX IV Laboratory. *urn:issn:1600-5775*, *28*, 1935–1947 , publisher = International Union of Crystallography. https://doi.org/10.1107/S1600577521008213

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Karlfeldt Fedje, K., Staničić, I., Nilsson, C., Bergman, F., Rui, H. M., Johansson, I., Sala, S., & Rissler, J. (2025). Speciation of Zn in WtE fly ash residues after wet treatment. *Waste Management*, *207*, 115112. https://doi.org/10.1016/j.wasman.2025.115112

Lee, D., & Seung, H. S. (2000). *Algorithms for Non-negative Matrix Factorization* https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791. https://doi.org/10.1038/44565

McInnes, L., & Healy, J. (2017, 18–21 Nov. 2017). Accelerated Hierarchical Density Based Clustering. 2017 IEEE International Conference on Data Mining Workshops (ICDMW),

Muehe, E. M., Adaktylou, I. J., Obst, M., Zeitvogel, F., Behrens, S., Planer-Friedrich, B., Kraemer, U., & Kappler, A. (2013). Organic Carbon and Reducing Conditions Lead to Cadmium Immobilization by Secondary Fe Mineral Formation in a pH-Neutral Soil. *Environmental science & technology*, *47*(23), 13430–13439. https://doi.org/10.1021/es403438n

Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, *5*(2), 111–126. https://doi.org/https://doi.org/10.1002/env.3170050203

Rissler, J., Fedje, K. K., Klementiev, K., Ebin, B., Nilsson, C., Rui, H. M., Klufthaugen, T. M., Sala, S., & Johansson, I. (2024). Zinc speciation in fly ash from MSWI using XAS - novel insights and implications. *Journal of Hazardous Materials*, *477*, 135203. https://doi.org/10.1016/j.jhazmat.2024.135203

Rissler, J., Klementiev, K., Dahl, J., Steenari, B. M., & Mar, E. (2020). Identification and Quantification of Chemical Forms of Cu and Zn in MSWI Ashes Using XANES. *Energy and Fuels*, *34*(11), 14505–14514 https://doi.org/10.1021/acs.energyfuels.0c02226

Rosipal, R., & Krämer, N. (2006, 2006//). Overview and Recent Advances in Partial Least Squares. Subspace, Latent Structure and Feature Selection, Berlin, Heidelberg.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. https://doi.org/https://doi.org/10.1016/S0169-7439(01)00155-1

NANO-XRF ANALYSIS FOR METAL SPECIATION IN ASH

Metals in waste incineration fly ash pose a challenge as their presence often leads to the ash being classified as hazardous waste. Yet these metals also represent valuable resources for recycling. Recovering metals from ash reduces the need for mining, improving resource utilization, and lowers the risk of metals leaching into the environment. Ash residues with low or non-soluble content of metals can be used in construction materials. Also in this context, the chemical form of the metals in the ash is essential as it governs solubility.

Analysis of nano-XRF elemental maps can provide important insights about which elements often occur colocalized on a sub-particle scale, thus helping to suggest possible chemical forms present in the fly ash. Analyzing large amounts of nano-XRF data comes with several challenges, such as collinearity of elemental concentrations and particles with mixed phases. Within this project several different statistical approaches for analysis of nano-XRF data were identified and evaluated. The overall aim is to explain which elements are often colocalized with transition elements such as Zn, an element important both from environmental leaching and extraction viewpoint, in fly ash particles.

The methods evaluated include pixel selection based on correlation trends, cluster analysis, and factor analysis. The evaluated methods have their pros and cons, whereof some could explain parts of the elemental behavior of Zn in the fly ash particles. But none could handle the full complexity of the ash data. We therefore suggest alternative options to be explored in future analysis efforts.

A new step in energy research

The research company Energiforsk initiates, coordinates, and conducts energy research and analyses, as well as communicates knowledge in favor of a robust and sustainable energy system. We are a politically neutral limited company that reinvests our profit in more research. Our owners are industry organisations Swedenergy and the Swedish Gas Association, the Swedish TSO Svenska kraftnät, and the gas and energy company Nordion Energi.

Energiforsk