SCRAP COMPOSITION ESTIMATION IN EAF AND BOF: IDENTIFIABILITY, CONDITIONING, AND THE KALMAN FILTER

[1]Yiqing ZHOU, [2]Karsten NAERT, [1]Dirk NUYENS

[1]KU Leuven, Department of Computer Science, Leuven, Belgium, EU, yiqing.zhou@kuleuven.be, dirk.nuyens@kuleuven.be

[2]ArcelorMittal of Belgium, Steel Shop Department, Gent, Belgium, EU, karsten.naert@arcelormittal.com

https://doi.org/10.37904/metal.2026.5230

Abstract

Estimating the elemental composition of scrap types used in Basic Oxygen Furnaces (BOF) and Electric Arc Furnaces (EAF) is fundamentally an ill-conditioned inverse problem. The composition of several scrap types must be recovered from a single scalar mass-balance measurement per heat, and the quality of any estimate depends critically on the condition number of the observation matrix formed by the scrap input masses. This paper systematically analyses how scrap-type usage frequency, charging diversity, and window length affect the condition number and explains why poor conditioning causes windowed least-squares methods to fail for rarely used scrap types or fixed charging recipes. We formulate scrap composition estimation as a linear Gaussian state-space model and show how the Kalman filter mitigates ill-conditioning and yields conservative and physically plausible estimates. We show on synthetic BOF data that the Kalman filter tracks time-varying composition more robustly than windowed non-negative least squares, particularly when the observation matrix is near-singular. The resulting scrap composition estimates can subsequently be used to support prediction of the elemental composition of the produced steel.

Keywords: EAF and BOF steelmaking, scrap composition, state-space model, Kalman filter

1. INTRODUCTION

Steel production increasingly relies on scrap metal as a raw material input [1,2]. Scrap is attractive both economically and environmentally: since EAF relies entirely on scrap while BOF combines scrap with hot metal, both routes reduce the need for virgin ore extraction and significantly decrease carbon emissions compared to primary steelmaking routes [3]. However, scrap is inherently heterogeneous and carries different levels of tramp elements. Many tramp elements, including Cu and Ni, cannot be effectively removed during conventional steelmaking, and excessive concentrations can degrade mechanical properties and surface quality [4,5]. Accurate knowledge of scrap composition is therefore essential for process control and maximizing the safe use of recycled material.

Existing approaches to scrap composition estimation share a common limitation: they produce only long-run averages rather than time-varying estimates [6-8]. Reference [6] formulates the problem using a mass balance and compares two estimators over a window of heats: non-negative least squares and maximum likelihood estimation. References [7,8] avoid the explicit mass-balance formulation and instead use regression models to predict steel composition from process variables: partial least squares in [7] and a nonlinear XGBoost model with SHAP attribution in [8]. The scrap composition can then be recovered by feeding hypothetical single-scrap-type heat inputs through the trained model. However, all these indirect approaches are affected by identifiability and conditioning problems. When scrap types are used rarely, or when several scrap types are consistently charged in fixed ratios, their individual compositions cannot be estimated reliably using windowed least-squares methods. This issue is discussed in Section 2.1. A separate line of work bypasses indirect

inference altogether by characterizing scrap composition through random sampling and optical emission spectroscopy of physical scrap samples [10]. This provides direct compositional measurements but requires substantial additional instrumentation and laboratory effort beyond what is collected during production.

Rather than solving a static linear system over a window of heats or requiring additional measurements, we estimate scrap composition recursively using a state-space model and the Kalman filter. We introduced this state-space framework and carried out a sensitivity analysis in [11]. The model captures the time-varying nature of scrap composition and provides uncertainty estimates for each heat, using only data already collected during standard production. Building on this foundation, the present paper analyses how conditioning affects windowed estimates and compares windowed NNLS with the Kalman filter using synthetic BOF data. We further demonstrate the connection between the Kalman gain, the observation matrix, and the identifiability of individual scrap-type compositions.

2. MATHEMATICAL MODELS

This section formulates scrap composition estimation in two ways. Section 2.1 presents the windowed inverse problem and discusses the role of conditioning in estimation quality. Sections 2.2 and 2.3 introduce the state-space formulation and the Kalman filter and explain how prior information stabilizes recursive estimation when the observations are weakly informative.

2.1 The inverse problem and its condition number

For a tramp element that is not transferred to the slag or gas phase, such as Cu or Ni, the mass balance equation at heat t in the BOF or EAF process is given by

$$\overrightarrow{m_t} \cdot \overrightarrow{\alpha_t} = y_t, y_t := m_{steel,t} f_{steel,t} - m_{hm,t} f_{hm,t}, \tag{1}$$

where is the scrap-input mass vector, is the vector of elemental fractions to be estimated, and is the tramp-element mass from scrap computed using equation (1). For the EAF case, the hot-metal mass is set to zero. Stacking N_w consecutive heats gives the linear system

$$A\vec{x} \approx \vec{b}, A = \begin{bmatrix} \vec{m}_t^T \\ \vdots \\ \vec{m}_{t+N_w-1}^T \end{bmatrix} \in R^{N_w \times N_s}, \vec{b} = \begin{bmatrix} y_t \\ \vdots \\ y_{t+N_w-1} \end{bmatrix}, \tag{2}$$

where $\vec{x}$ represents the scrap composition vector, assumed to remain constant over the selected window.

We assume the observation noise on y_t is normally distributed with zero mean and standard deviation σ. Then the ordinary least squares (OLS) estimator and its covariance matrix are given by, see [9],

$$\hat{\vec{x}}_{OLS} = (A^T A)^{-1} A^T \vec{b}, Cov\hat{\vec{x}}_{OLS} = (A^T A)^{-1} \sigma^2. \tag{3}$$

The condition number $\kappa(A^T A)$ governs the sensitivity of the estimated composition to observation noise. Note that if the observation matrix A is of full column rank, then $\kappa(A^T A) = \kappa(A)^2$.

Elemental fractions must lie in [0,1], which is not guaranteed by OLS. Using windowed Non-Negative Least Squares (NNLS), we can constrain them to be at least non-negative [12]. However, the constraint introduces bias [12], provides no direct uncertainty quantification, and does not resolve the underlying ill-conditioning.

Common real-world scenarios from the steelmaking industry lead to a large condition number. (1) Infrequent usage. If a scrap type is used at low mass in only a few heats within the window, the corresponding column of A has small norm, which increases the condition number of $(A^T A)$ and can inflate the variance of the corresponding component of $\hat{\vec{x}}$. (2) Fixed recipe. If two or more scrap types are always charged together in a fixed mass ratio, their columns in A are nearly linearly dependent. Consequently, $A^T A$ is nearly singular, and

the individual fractions cannot be separated reliably, regardless of the number of observed heats. (3) Small window size. A window of size $N_w < N_s$ will make A rank-deficient.

2.2 State-space formulation

Because scrap composition varies over time, we formulate the problem as a state-space model [11] that tracks $\vec{\alpha}_t$ heat by heat:

$$\vec{\alpha}_{t+1} = (1-\gamma)\vec{\alpha}_t + \gamma\vec{\eta}_t, \tag{4}$$

$$y_t = \vec{m}_t \cdot \vec{\alpha}_t + \varepsilon_t, \tag{5}$$

where $t = 1,2,\ldots$ indexes heats, $0 < \gamma < 1$ controls composition changes, $\vec{\eta}_t \in R^{N_s}$ is Beta-distributed process noise ensuring fractions remain in [0,1], and $\varepsilon_t \sim N(0, H_t)$ is Gaussian observation noise.

2.3 Gaussian approximation and Kalman filter

Approximating $\vec{\eta}_t$ by the Gaussian distribution $N(\vec{q}, Q)$ simplifies equations (4) and (5) to a linear Gaussian state-space model, for which the Kalman filter provides the minimum-mean-squared error estimator [13]. The Beta distribution can be approximated by Gaussian distribution when the variance is moderate, which holds for most scrap types in practice. The corresponding algorithm for our model can be found in Figure 1.

KalmanStep $((\vec{m}_t, y_t); (\vec{a}_t, P_t); \gamma, \vec{q}, Q, H_t)$

$\backslash\backslash\ (y_t = m_{\text{steel},t} f_{\text{steel},X,t} - m_{\text{hm},t} f_{\text{hm},X,t})$
$Z_t = \vec{m}_t^{\top}$

Update (incorporate measurement)
$K_t = P_t Z_t^{\top} (Z_t P_t Z_t^{\top} + H_t)^{-1}$
$\vec{a}_{t|t} = \vec{a}_t + K_t (y_t - Z_t \vec{a}_t)$
$P_{t|t} = (I - K_t Z_t) P_t$

Predict (propogate to next heat)
$\vec{a}_{t+1} = (1-\gamma)\,\vec{a}_{t|t} + \gamma\,\vec{q}$
$P_{t+1} = (1-\gamma)^2 P_{t|t} + \gamma^2 Q$

Figure 1 Kalman filter update and prediction steps

Here K_t is the **Kalman gain computed from the prior covariance** P_t, the observation vector Z_t, and the observation noise variance H_t, which balances between the prior estimate and the new measurement. Assuming a diagonal prior covariance matrix since the compositions of different scrap types are independent, the i-th component of K_t can be written as $K_{t,i} = P_{t,ii}\, m_{t,i} / (Z_t P_t Z_t^T + H_t)$, where $m_{t,i}$ is the mass of scrap type i at heat t and $P_{t,ii}$ is the current marginal variance of the estimate for scrap type .

It is worth contrasting this behavior with the condition number analysis of Section 2.1. The Kalman filter never solves an inverse problem of the form $A\vec{x} \approx \vec{b}$ at each heat, so the condition number that governs the windowed estimators has no direct analogue here. Instead of inverting a concatenated observation matrix, the filter maintains a posterior distribution over $\vec{\alpha}_t$ that is updated per heat. The problems identified above can now be reconsidered. (1) Infrequent usage. When scrap type i is not used at heat t, then $m_{t,i} = 0$ and $K_{t,i} = 0$, leaving the corresponding component of $\vec{a}_{t\vee t}$ in the update step unchanged. The prediction step pulls the estimate toward the prior mean $\vec{q}$, providing a conservative and physically plausible result rather than an unbounded excursion. (2) Fixed recipe. When two scrap types i and j are always charged in a fixed mass ratio $m_{t,j} = r m_{t,i}$, the Kalman gain components are $K_{t,i} = P_{t,ii}\, m_{t,i} / (Z_t P_t Z_t^T + H_t)$ and $K_{t,j} = P_{t,jj} r\, m_{t,i} / (Z_t P_t Z_t^T + H_t)$, which are tied together by the same ratio r. The innovation $y_t - Z_t \overline{\alpha_t}$ is distributed between $\vec{a}_{t\vee t,i}$ and $\vec{a}_{t\vee t,j}$ according to $P_{t,ii}$ and $P_{t,jj} r$. Hence the individual estimates remain close to their priors, which is a controlled outcome rather than the unbounded excursions seen with windowed least squares. (3) Small window size.

The filter has no window but updates the scrap composition estimates for each heat, where γ describes the time correlation of the states between two heats. It is important to note that while the Kalman filter regularizes ill-conditioning, it does not restore identifiability that is structurally absent from the observations.

3. NUMERICAL EXPERIMENTS

We generate synthetic BOF data from equations (4) and (5) and compare Kalman filter and NNLS estimates of scrap type Cu fractions with the true value.

3.1 Synthetic data generation

To simulate a realistic scenario, we generate the scrap composition such that it contains clean scrap types and dirty scrap types. The scrap composition might also suddenly change based on different origins, which we will simulate by having a sudden jump in the scrap composition. We simulate $N_s = 5$ scrap types over 1000 heats. The true Cu fractions $\vec{\alpha}_t$ are generated according to equation (4) with $\gamma = 0.01$, meaning composition evolves slowly between heats. The initial Cu fractions are set to (100, 100, 500, 800, 1300) ppm respectively. Two abrupt shifts are introduced: the Cu fraction of scrap type 2 jumps from 100 to 400 ppm at heat 400, and the Cu fraction of scrap type 4 increases from 800 to 1000 ppm at heat 700. The standard deviations of $\vec{\eta}_t$ are set to (100, 200, 500, 1000, 1000) ppm, reflecting the greater compositional variability of scrap types with higher impurity levels. In a real steelmaking process, often clean scrap types dominate the charge, while higher-impurity types are used sparingly. To mimic this, the total scrap mass per heat follows a Gaussian distribution with a mean of 50 t and a standard deviation of 5 t, and the mass share of each scrap type is drawn from a Dirichlet distribution, where the first 500 heats have a mixing ratio of (0.50, 0.40, 0.09, 0.005, 0.005), and the last 500 heats have a mixing ratio of (0.50, 0.40, 0.099, 0.001, 0.0). To further simulate infrequent usage, any generated scrap mass smaller than 1.0×10^{-5} t is set to zero, ensuring that rarely used scrap types are absent from certain heats rather than present with a negligibly small mass.

The remaining BOF variables, m_{hm} and f_{hm}, are drawn from uniform distributions, with ranges summarized in **Table 1**. We compute the fraction of Cu ending up in the steel by

$$f_{steel} = \left(\vec{m}_{scrap} \cdot \vec{f}_{scrap} + m_{hm} f_{hm}\right) / m_{steel}. \quad (6)$$

The observation y_t is generated by adding Gaussian noise with variance H_t to be 3 kg^2 to $\overrightarrow{m_t} \cdot \overrightarrow{\alpha_t}$.

Table 1 Physical meaning and distributions of the simulated BOF variables

Variable	Physical meaning	Range of uniform distribution
m_{hm}	Mass of hot metal	[200, 300] t
f_{hm}	Cu fraction of hot metal	[50, 100] ppm
m_{steel}	Mass of steel	[330, 360] t
f_{steel}	Cu fraction of steel	to be calculated by Equation (6)

3.2 Comparison of Kalman filter and windowed NNLS

The Kalman filter results are compared against windowed NNLS. Both methods are given their most favourable hyperparameters. The hyperparameters for the Kalman filter are chosen to be the best possible values based on the generating step. The window size for NNLS was selected by a grid search over the candidate values (10, 20, 30, 40, 50, 60, 100, 200). For each candidate, the composition of scrap type 2 estimated by NNLS was compared with its generated true value, and the window size minimizing the root-mean-square error over all heats after the initial 200 heats. This results in a window size of 30 heats, giving good tracking for frequently

charged scrap types but worse behavior for less frequently used types. Results for scrap types 2 and 5 are shown in **Figure 2**, chosen as representative examples of a frequently used scrap type undergoing an abrupt composition change and a rarely used scrap type.

As can be seen in the left plot of **Figure 2**, when a scrap type is used regularly, both the Kalman filter and NNLS track the true composition well during steady periods. However, when a sudden jump occurs, such as the imposed composition change at heat 400, NNLS requires several heats to adapt, with the lag determined by the window size. A smaller window reduces the lag but increases instability, since fewer observations worsen the condition number of the NNLS matrix and amplify estimation uncertainty. The Kalman filter also requires some heats to adjust, but the Kalman gain naturally weighs recent observations more heavily when the prediction error of the observation is large.

As a second observation, the right panel of **Figure 2** shows that when a scrap type is used rarely or in small quantities, NNLS produces implausible estimates, including large excursions far from physically reasonable values. This occurs because the corresponding column of the NNLS matrix is near zero, making the system poorly conditioned and the estimate unreliable. The Kalman filter behaves more robustly in this regime: when measurements carry little information about a particular scrap type, the update step contributes little and the estimate reverts toward the prior mean $\vec{q}$, yielding a conservative and physically plausible result.

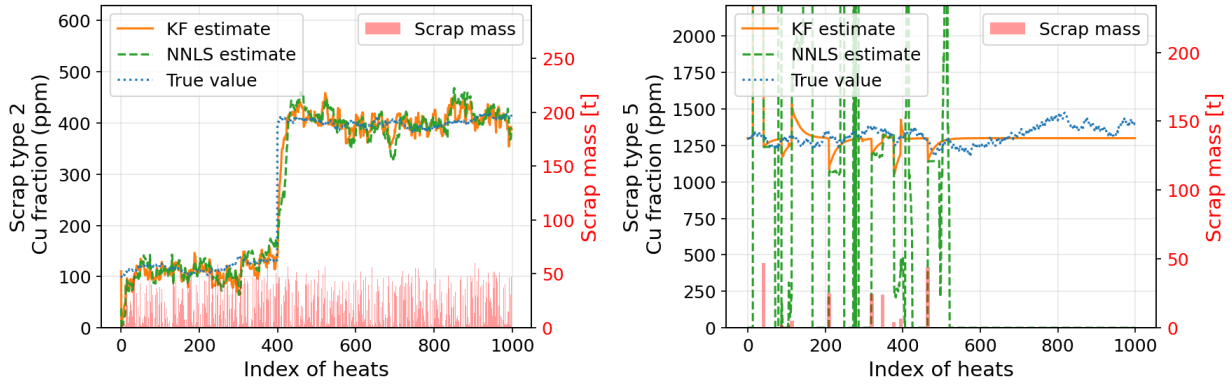

Figure 2 True and estimated Cu fractions and charged masses for scrap type 2 (left) and scrap type 5 (right)

4. CONCLUSION

Scrap composition estimation is limited by how much information the charging data carry about each individual scrap type, and this limitation depends on usage frequency and charging diversity. The two methods here presented expose it differently. In windowed NNLS it appears as a large condition number of the observation matrix: infrequent usage, fixed charging recipes, and small window sizes all make the estimate highly sensitive. The synthetic experiments confirm this, with NNLS producing large excursions far from physically reasonable values for the rarely used scrap type. There is also a trade-off in the window size: after an abrupt composition change, NNLS adapts only with a lag set by the window, while a smaller window shortens the lag but worsens the condition number. The Kalman filter never inverts a concatenated observation matrix, so its condition number has no counterpart there. The same lack of information instead enters through the Kalman gain, such that an uncharged scrap type leaves the corresponding estimate unchanged in the update step. The filter therefore gives a conservative and physically plausible result, together with a per-heat uncertainty estimation that windowed NNLS does not provide. It does not restore identifiability, so these estimates remain prior-dependent. The resulting scrap composition estimates can subsequently support prediction of the elemental composition of the produced steel, as demonstrated by the numerical experiments.

ACKNOWLEDGEMENTS

The research leading to these results has been performed within the CAESAR project and received funding from the European Commission's Horizon Europe Programme under grant agreement No. 101058520.

REFERENCES

[1] HAY, T., VISURI, V. V., AULA, M., ECHTERHOF, T. A review of mathematical process models for the electric arc furnace process. Steel Research International. 2021, vol. 92, pp. 2000395.

[2] DE VOS, L., BELLEMANS, I., VERCRUYSSEN, C., VERBEKEN, K. Basic Oxygen Furnace: assessment of recent physicochemical models. Metallurgical and Materials Transactions B. 2019, vol. 50, pp. 2647-2666.

[3] WU, H., CHENG, J., ZHONG, H., JIN, F. Recent progress in iron and steel industry decarbonization strategies: industrial advancements and challenges. Environmental Science and Pollution Research. 2025, vol. 32, pp. 4445-4471.

[4] DAEHN, K. E., CABRERA SERRENHO, A., ALLWOOD, J. M. How will copper contamination constrain future global steel recycling? Environmental Science & Technology. 2017, vol. 51, pp. 6599-6606.

[5] CEJKA, J., SAMMER, B., GRUBER, I., KLÖSCH, G., MICHELIE, S. K. Influence of tramp elements on surface properties of liquid medium-carbon steels. Steel Research International. 2024, vol. 95, pp. 2300715.

[6] ARZPEYMA, N., ALAM, M., GYLLENRAM, R., JÖNSSON, P. G. Model development to study uncertainties in electric arc furnace plants to improve their economic and environmental performance. Metals. 2021, vol. 11, pp. 892.

[7] SANDBERG, E., LENNOX, B., UNDVALL, P. Scrap management by statistical evaluation of EAF process data. Control Engineering Practice. 2007, vol. 15, pp. 1063-1075.

[8] SCHÄFER, M., FALTINGS, U., GLASER, B. Artificial Intelligence-based back-calculation model for scrap compiling optimization. Engineering Applications of Artificial Intelligence. 2026, vol. 167, pp. 113809.

[9] MONTGOMERY, D. C., PECK, E. A., VINING, G. G. Introduction to Linear Regression Analysis. Hoboken: John Wiley & Sons, 2021.

[10] GAUFFIN, A., TILLIANDER, A., JÖNSSON, P. G. Alloy content in steel scrap by use of random sampling analysis and its impact on the Electric arc furnace. In: Proceedings of the Shechtman International Symposium. Cancun: Taylor & Francis. 2014, vol. 29, pp. 6.

[11] ZHOU, Y., NAERT, K., NUYENS, D. Sensitivity Analysis of State Space Models for Scrap Composition Estimation in EAF and BOF. In: Journal of Physics: Conference Series. Ljubljana: IOP. 2025, vol. 3050, pp. 012001.

[12] LAWSON, C. L., HANSON, R. J. Solving least squares problems. Philadelphia: Society for Industrial and Applied Mathematics, 1995.

[13] DURBIN, J., KOOPMAN, S. J. Time Series Analysis by State Space Methods. Oxford: Oxford University Press, 2012.