

<https://doi.org/10.31891/csit-2026-3-21>**Vitalii TRON**

Candidate of Technical Sciences
Associate Professor, Department of
Automation, Computer Science and
Technology

Kyryvyi Rih National University

<https://orcid.org/0000-0002-6149-5794>

e-mail: vtron@knu.edu.ua

Received: 10/07/2026

Accepted: 30/08/2026

Published: 30/09/2026

© Copyright
2026 by the author(s)



This is an Open Access article distributed
under the terms of the [Creative Commons
CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

UDC 622.7:519.246

SEMIPARAMETRIC IDENTIFICATION AND UNCERTAINTY QUANTIFICATION OF BENEFICIATION UNIT SEPARATION CHARACTERISTICS

Identifying the separation characteristics of iron ore beneficiation units from process data provides the model and uncertainty information required for model predictive control. Parametric models, including the Rosin–Rammler and sigmoid models, have fixed analytical forms that may not describe the characteristics of individual units with sufficient accuracy. Point estimation without a dedicated statistical procedure does not quantify predictive uncertainty, while purely nonparametric estimates may be unstable when data are limited. This study developed a unified semiparametric identification method in which a physically motivated parametric prior mean function was augmented with a correction estimated by Gaussian process regression from empirical deviations of the observations. The same estimation procedure was used for model fitting, cross-validation, and sensitivity analysis. A semiparametric Nadaraya–Watson estimator constructed from deviations from the prior mean function served as a benchmark. Experimental recovery was calculated using a common mass-balance definition. The ratio between product mass flow rates was either obtained from measurements or estimated from the size-by-size balance. For the classifier, this ratio was well constrained by the data, with a 95% bootstrap interval of [1.95, 2.42]. The proposed semiparametric model reduces the within-sample root-mean-square error by approximately 82% relative to the Rosin–Rammler model and outperforms the stand-alone Gaussian process model in leave-one-out cross-validation, with coefficients of determination of 0.970 and 0.945, respectively. The mass-flow-rate ratio was re-estimated in every cross-validation iteration using only the training size classes. For the hydrocyclone, the ratio was estimated with a 95% bootstrap interval of [0.96, 2.56], and the proposed model retains the highest leave-one-out coefficient of determination, approximately 0.81, throughout this interval. Results for magnetic separators with measured mass flow rates confirm that the argument of a separation characteristic must reflect the physical separation mechanism. Negative leave-one-out coefficients of determination obtained when the separation characteristic was modelled as a function of particle size demonstrate that particle size is an unsuitable argument and justify representing the characteristic as a function of iron content, which more directly reflects the magnetic separation mechanism. Monte Carlo simulation under the specified noise model was used to quantify sensitivity; for the classifier, the mean relative width of the uncertainty band is approximately 14%.

Keywords: *semiparametric identification; separation characteristic; Gaussian process regression; kernel correction; mass balance; cross-validation; uncertainty quantification; reproducibility; iron ore beneficiation.*

Problem statement

The performance of model predictive control (MPC) for iron ore beneficiation processes depends substantially on the accuracy of process-unit models. A key component of these models is the separation characteristic (partition curve), which gives the fraction of a given particle class reporting to a specified product as a function of particle size d for classifiers and hydrocyclones or iron content Fe for magnetic separators. Separation characteristics are conventionally approximated using parametric models with a fixed analytical form. Their use in adaptive MPC, however, involves several limitations.

First, a fixed analytical form does not always represent nonstandard characteristics: a parametric approximation may yield a negative coefficient of determination for units with recirculation and for magnetic separators operating in the saturation region. Second, when the mineralogical and processing properties of the feed change, complete least-squares re-

identification of all unit parameters in a process line can be computationally demanding. If the models of numerous units are re-identified sequentially, the total computation time may approach the interval between such changes. Third, parametric point estimation without a dedicated statistical procedure does not quantify parameter or predictive uncertainty. Conventional intervals may be unreliable for dependent, heteroskedastic, or small samples, although parametric regression generally supports covariance, confidence, and prediction intervals. Robust MPC requires a consistent quantitative measure of uncertainty in the identified separation characteristic. Fourth, strong autocorrelation of regression residuals complicates identification from process time series. Under these conditions, standard-error estimates computed under classical least-squares assumptions may be biased, and the corresponding confidence intervals may not attain their nominal coverage.

These considerations motivate a unified method for identifying separation characteristics that combines the physical interpretability of parametric models with the flexibility of nonparametric estimators and quantifies uncertainty in accordance with the structure of the available data.

Literature review

The parametric model for a separation characteristic is selected according to the physical separation mechanism and the particle property that governs reporting to the corresponding product. For classifiers and hydrocyclones, this property is particle size d : the proportion of material reporting to the overflow generally decreases as particle size increases. Size-based partition curves and their application to the modelling of beneficiation, grinding, and classification processes are discussed in [1, 2]. A decreasing Rosin–Rammmler model is suitable as a physically interpretable prior mean function for $\varepsilon(d)$ [3]. Its parameter d_{50} is the median cut size, whereas n characterizes separation sharpness.

For magnetic separators, the determining factor is the magnetic susceptibility of particles, which is related to their mineralogical composition and iron content [1]. The process outcome can therefore be described by the recovery of iron to the magnetic product, $\eta(\text{Fe})$. Recovery increases with iron content and approaches a limiting level, which supports the use of an increasing sigmoid function. In this model, $\eta_{\max}$ specifies the limiting recovery, Fe_{50} is the iron content at which recovery equals one-half of this limit, and k characterizes the rate of change in the transition region.

The Rosin–Rammmler and sigmoid models are thus physically interpretable prior mean functions associated with different separation mechanisms. Their fixed analytical forms represent the principal trend and physically plausible limiting behaviour of the characteristic, although they may not capture local deviations caused by changes in feed properties and equipment operating conditions with sufficient accuracy.

Addressing this limitation has motivated identification methods that infer relationships directly from data. A systematic review of machine-learning applications in mineral processing is provided in [4], with primary emphasis on flotation, machine vision, and process-equipment diagnostics. The identification of separation characteristics using a physically motivated prior mean function together with consistent uncertainty quantification is not considered separately in that review.

Nonparametric and kernel methods are promising for reconstructing nonlinear separation characteristics from limited samples. These methods include Gaussian process regression (GPR) [5] and the Nadaraya–Watson (NW) estimator [6, 7]. They adapt the shape of the estimated relationship to the structure of the observed data. Under the corresponding modelling assumptions, GPR also provides a posterior uncertainty estimate. With limited observations, however, nonparametric estimates may be unstable, and selecting the kernel bandwidth requires a dedicated procedure [8]. Kernel estimators for identifying mineral-processing units have also been developed in the author’s previous studies [9, 10].

Hybrid and semiparametric structures combine the physical interpretability of parametric models with the flexibility of data-driven methods. In these structures, deviations of the observations from a prior mean function are used to estimate a correction term. In [11], an additive GPR correction to the nominal dynamics was used to construct probabilistic MPC constraints. Bootstrap methods are used to quantify parameter and predictive uncertainty [12], while heteroskedasticity- and autocorrelation-consistent (HAC) covariance estimators [13] and the stationary block bootstrap [14] are appropriate when process time series are autocorrelated.

The reviewed studies indicate that parametric separation characteristics, kernel correction methods, and uncertainty-quantification procedures have mainly been developed as separate methodological components. Their coordinated use for identifying the separation characteristics of mineral-processing units remains insufficiently investigated. This gap defines the objective of the present study.

Research objective

The objective is to develop a unified method for the semiparametric identification of separation characteristics in iron ore beneficiation. The method combines a physically motivated parametric prior mean function with a nonparametric kernel correction estimated from empirical deviations of observations from that function, quantifies uncertainty according to the structure of the available data, and empirically assesses its applicability and limitations using data from different types of process units.

Achieving this objective requires defining physically interpretable prior mean functions, constructing the kernel correction term, selecting uncertainty-quantification procedures, and evaluating the predictive performance of the resulting models.

Methods and results

Definition of a separation characteristic and parametric models

The separation characteristic of a size-separation unit specifies the mass fraction of feed material in size class d that reports to the selected product. For classifiers and hydrocyclones, product 1 is defined as the overflow (fine product): $\varepsilon(d) \rightarrow 1$ for fine particles ($d \rightarrow 0$), whereas $\varepsilon(d) \rightarrow 0$ for coarse particles ($d \rightarrow \infty$) reporting to the underflow. In the general case, the experimental recovery of the i th size class to product 1 is determined from the mass balance as

$$\varepsilon_i = \frac{Q_1 \gamma_{1,i}}{Q_f \gamma_{f,i}}, \quad (1)$$

where Q_f is the solids mass flow rate in the feed; Q_1 is the mass flow rate of the product for which recovery is calculated; and $\gamma_{f,i}$ and $\gamma_{1,i}$ are the mass fractions of the i th size class in the feed and product 1, respectively. For two-product separation, $Q_f = Q_1 + Q_2$, where Q_2 is the mass flow rate of the second product stream and $\gamma_{2,i}$ is the mass fraction of the i th size class in product 2; the recoveries of the i th class to the two products sum to unity. The definition in (1), with $\varepsilon \rightarrow 1$ for fine classes reporting to the overflow, corresponds to the decreasing Rosin–Rammler (R–R) model [3] written with the $\ln 2$ factor:

$$\varepsilon(d) = \exp\left(-\ln 2 \left(\frac{d}{d_{50}}\right)^n\right), \quad (2)$$

for which $\varepsilon(d_{50}) = 0.5$. Thus, d_{50} is the actual cut size, that is, the median size at which the mass of the class is divided equally between the overflow and underflow, and n is the separation-sharpness exponent. In the form without $\ln 2$, the parameter represents the characteristic scale d_c at which $\varepsilon = e^{-1}$. The factor $\ln 2$ gives d_{50} its median interpretation, which is standard for hydrocyclones. The parameters (d_{50}, n) are re-estimated in every identification procedure, including every cross-validation iteration.

For magnetic separators, the principal separation factor is magnetic susceptibility, which is correlated with iron content. The recovery of iron to the magnetic product as a function of Fe content is therefore approximated by the increasing sigmoid model [1]:

$$\eta(\text{Fe}) = \frac{\eta_{\max}}{1 + \exp(-k(\text{Fe} - \text{Fe}_{50}))}, \quad (3)$$

where $\eta_{\max}$ is the maximum recovery, k is the selectivity parameter, and Fe_{50} is the iron content at which recovery equals one-half of the maximum. Model (2) yields $R^2 > 0.98$ within the sample for the considered size-separation units. The sigmoid model (3) reproduces the principal trend in the available magnetic-separator data ($R^2 \approx 0.71$). The empirical mean is used as the prior mean function when describing $\varepsilon(d)$ for the deslimmer and magnetic separators; for the magnetic separators, the characteristic $\eta(\text{Fe})$, whose argument is related to particle magnetic properties, is also considered.

The accuracy of separation-characteristic identification for subsequent use in MPC is improved using a semiparametric approach in which a physically interpretable prior mean function is augmented with a data-estimated correction term.

Semiparametric model with kernel correction

A semiparametric model is proposed to combine the advantages of physically interpretable parametric models and nonparametric estimators. Its prediction is the sum of a prior mean function $m(\xi)$ and a nonparametric kernel correction $\delta(\xi)$ that describes the systematic component of the deviations of observed values from this function. The prior mean function is a physically motivated baseline relationship that defines the expected shape of the separation characteristic before refinement using observations. The resulting structure is

$$\hat{\varepsilon}(\xi) = m(\xi) + \delta(\xi), \quad r_i = \varepsilon_i - m(\xi_i), \quad (4)$$

where ξ is the separation variable ($\xi = d$ for size separation and $\xi = \text{Fe}$ for magnetic separation); $m(\xi)$ is given by the Rosin–Rammler model (2) for size separation, the sigmoid model (3) for magnetic separation by iron content, or the empirical mean $\bar{\varepsilon}$ for units operating in the saturation region; r_i is the regression residual, defined as the empirical deviation of observation ε_i from the prior-function value $m(\xi_i)$; and $\delta(\xi)$ estimates the systematic component of these deviations.

In the primary implementation, the correction $\delta(\xi)$ is estimated using GPR. As a benchmark, the systematic component of the same empirical deviations is estimated using the NW estimator [6]:

$$\hat{\delta}_{\text{NW}}(\xi) = \frac{\sum_{i=1}^n K_h(\xi - \xi_i) r_i}{\sum_{i=1}^n K_h(\xi - \xi_i)}, \quad (5)$$

where K_h is a Gaussian kernel with bandwidth h . Normalization by the sum of the kernel weights makes this estimate a locally weighted mean of the empirical deviations r_i , but it does not quantify uncertainty. The primary implementation uses GPR [5] with a radial basis function kernel:

$$\hat{\delta}_{\text{GPR}}(\xi^*) = k(\xi^*, D)^\top [K(D, D) + \sigma_n^2 I]^{-1} r, \quad (6)$$

where $K(D, D)$ is the kernel Gram matrix at observation points D , σ_n^2 is the noise variance, and r is the vector of regression residuals defined as empirical deviations from the prior mean function. The hyperparameters, namely the length scale and σ_n , are determined by maximizing the log marginal likelihood. The posterior mean in (6) has a kernel-ridge form and therefore does not generally coincide with the NW estimator in (5). It also provides the posterior

predictive variance used to construct uncertainty bands. The final prediction is constrained to the physical range $\hat{\varepsilon} \in [0,1]$. For the size-separation units, the same complete estimator—the Rosin–Rammler prior mean function with a GPR correction estimated from r_i , with all parameters re-estimated—is used for full-sample estimation, cross-validation, and Monte Carlo simulation.

The estimate in (4) interpolates between the parametric prior mean function and the nonparametric correction according to the number n_{tr} of informative points in the transition region, defined by $0.02 < \varepsilon_i < 0.98$ and $d_i > 10 \mu\text{m}$. When n_{tr} is small, regularization through a large σ_n and length scale limits the correction amplitude; the estimate consequently remains close to the parametric model and retains its physically plausible limiting behaviour. With sufficient n_{tr} , the correction becomes more flexible and captures nonstandard shapes. The kernel bandwidth of the semiparametric NW estimator is determined by leave-one-observation-out cross-validation; for $\varepsilon(d)$, each observation represents one size class:

$$h_{\text{opt}} = \arg \min_h \sum_{i=1}^n (r_i - \hat{\delta}_{-i}(\xi_i; h))^2. \quad (7)$$

At each iteration, the i th observation is temporarily removed, and the correction $\hat{\delta}_{-i}$ is estimated from the remaining $n - 1$ observations. The estimate at ξ_i is compared with the empirical deviation r_i . The optimal bandwidth h minimizes the sum of squared prediction errors over all successively held-out observations; the initial value is obtained using Silverman’s rule [8]. The GPR hyperparameters are re-estimated in each cross-validation iteration by maximizing the marginal likelihood.

The proposed structure therefore combines a physically interpretable prior mean function with an adaptive correction whose parameters are determined using a common estimation procedure. Quantitative measures of the accuracy and uncertainty of the resulting estimate are obtained in accordance with the structure of the available data.

Data-specific uncertainty quantification

Uncertainty-quantification procedures are selected according to the structure of the available data because no single procedure is equally suitable for all datasets.

The primary separation dataset was obtained from a single cross-sectional sieve-analysis survey conducted during plant trials and comprises nine size classes measured at up to nineteen sieve-sampling points. For such data, predictive performance is evaluated by leave-one-out cross-validation (7), while uncertainty in parametric-model parameters, such as Fe_{50} of the sigmoid, is evaluated by bootstrapping regression residuals [12]. Robustness to measurement errors is analysed by Monte Carlo simulation, with replicates generated under a specified noise model. The resulting band characterizes the sensitivity of the identified characteristic to the adopted noise model.

Process data from supervisory control and data acquisition (SCADA) systems are time series with strong autocorrelation and require different tools: blocked (rolling-window) time-series cross-validation, the Newey–West HAC covariance estimator [13], and the stationary block bootstrap [14] with a block length exceeding the autocorrelation horizon. The HAC correction accounts for autocorrelation and heteroskedasticity when estimating the covariance matrix; endogeneity, nonstationarity, and coefficient bias are assessed using separate diagnostic procedures.

The resulting point estimates and measures of dispersion provide an interface for supplying model and uncertainty information to local MPC optimization problems for constraint formulation and to a coordinator that exchanges uncertainty estimates among agents. Closed-loop MPC experiments form a subsequent stage in the quantitative evaluation of robustness and probabilistic-constraint satisfaction.

This set of procedures aligns the evaluation of predictive performance and uncertainty with the data type. It is applied when identifying the separation characteristics of the investigated process units.

Identification results for process units

The performance of the proposed GPR-based semiparametric model (4) is evaluated using data from plant-scale process tests at an ore beneficiation plant. It is compared with the Rosin–Rammler model, a stand-alone GPR model without a physically motivated prior mean, and a semiparametric NW benchmark in which the correction to the prior mean function is estimated from the empirical deviations r_i . For every model, experimental recovery is calculated using the common mass-balance definition in (1). For two-product separation, the product mass-flow-rate ratio is $r_Q = Q_2/Q_1$; this ratio is invariant under common proportional scaling of both mass flow rates.

The provenance of each r_Q value is recorded in a machine-readable artifact. For the classifier and hydrocyclone, product 1 is the overflow and product 2 is the underflow. When a feed sieve analysis is available, r_Q is estimated by least squares from the size-by-size balance $(1 + r_Q)\gamma_{f,i} = \gamma_{1,i} + r_Q\gamma_{2,i}$. For magnetic separators, product 1 is the magnetic product and product 2 is the nonmagnetic product, and r_Q is calculated from their measured mass flow rates. The classifier estimate $r_Q = 2.08$ is well constrained, with a 95 % bootstrap interval of [1.95, 2.42], a size-by-size mass-balance root-mean-square residual of 0.55 percentage points, and negligible sensitivity to the omission of any single size class.

For the classifier, where r_Q is well constrained and the transition region contains $n_{tr} = 7$ points, the proposed GPR-based semiparametric model yields within-sample values of $R^2 = 0.999$ and $\text{RMSE} = 0.008$. Its RMSE is approximately 82 % lower than that of the Rosin–Rammler model (0.043; Fig. 1). The stand-alone GPR model and the semiparametric NW benchmark reproduce the within-sample data more closely, with RMSE values of 0.005 and

0.004, respectively. The within-sample evaluation is supplemented by leave-one-out cross-validation of predictive performance. The proposed model then outperforms the stand-alone GPR model ($R_{\text{LOO}}^2 = 0.970$ versus 0.945; Table 1), indicating that the physically motivated prior mean benefits generalization.

The semiparametric NW benchmark constructed from the same deviations r_i yields a similar $R_{\text{LOO}}^2 = 0.976$. Given the available sample size, the gain in predictive performance is appropriately attributed to the combination of the prior mean function and the nonparametric correction. A larger sample would enable the performance gains attributable to the GPR and NW corrections to be distinguished statistically. GPR is selected as the primary implementation because it provides a posterior variance under the modelling assumptions.

For the hydrocyclone, r_Q is estimated with a 95 % bootstrap interval of [0.96, 2.56]. At $n_{\text{tr}} = 4$, the proposed GPR-based semiparametric model achieves the best predictive performance among the compared models: $R_{\text{LOO}}^2 = 0.81$, compared with 0.69 for Rosin–Rammler and 0.76 for the stand-alone GPR model. Sensitivity analysis shows that this advantage is maintained throughout the identified range of r_Q , supporting the robustness of the comparative result to uncertainty in the mass-flow-rate ratio. The resulting d_{50} range of approximately 16–30 μm quantifies the sensitivity of this parameter estimate to r_Q .

Results for magnetic separators with measured mass flow rates confirm the importance of selecting the argument of the separation characteristic according to the physical separation mechanism. For $\varepsilon(d)$, leave-one-out cross-validation gives a negative R_{LOO}^2 despite a close within-sample representation ($R^2 \approx 0.91$). This result supports the use of $\eta(\text{Fe})$ in (3), whose argument is directly related to particle magnetic properties.

Table 1.

Comparison of within-sample accuracy and predictive performance of the separation-characteristic models

Unit	r_Q provenance	n_{tr}	R–R		GPR		Semiparametric	
			R_{in}^2	R_{LOO}^2	R_{in}^2	R_{LOO}^2	R_{in}^2	R_{LOO}^2
Classifier	size-by-size balance	7	0.981	0.967	1.000	0.945	0.999	0.970
Hydrocyclone	size-by-size balance	4	0.983	0.690	0.988	0.762	0.984	0.809
Magnetic separator, $\varepsilon(d)$	measured flow rates	7	–	–	0.914	< 0	0.914	< 0

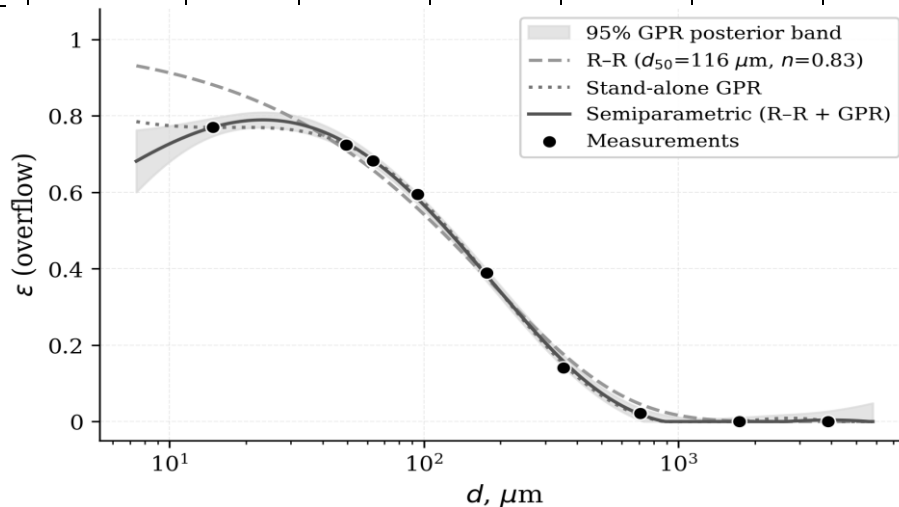


Fig. 1. Classifier separation characteristic

The calculations show that, for the classifier, the stand-alone GPR model and the semiparametric implementations with GPR and NW corrections yield RMSE values of 0.004–0.008, compared with 0.043 for the Rosin–Rammler model. For the hydrocyclone, the RMSE values of the compared models are 0.019–0.023 (Fig. 2).

The joint analysis of within-sample accuracy and cross-validation results therefore supports combining a physically motivated prior mean function with a kernel correction term. Cross-validation and measurement-error simulation further characterize model generalization and the robustness of the resulting estimates.

Validation of predictive performance and robustness

Predictive performance is evaluated by leave-one-out cross-validation using the same GPR-based semiparametric estimator, which combines the Rosin–Rammler prior mean function with a GPR correction estimated from r_i . At every iteration, the prior-mean parameters, GPR hyperparameters, and mass-flow-rate ratio r_Q are re-estimated using only the training size classes. Because ε_i depends on r_Q , estimating r_Q from the complete sample would indirectly leak information about the held-out size class into its target value. Re-estimating r_Q at every iteration excludes the held-out class from this calculation.

For the classifier, $R_{LOO}^2 = 0.970$ at $RMSE_{LOO} = 0.054$; without re-estimating r_Q , R_{LOO}^2 is slightly inflated to 0.974. Comparison with the within-sample value $RMSE = 0.008$ reveals a substantial difference between within-sample and cross-validation errors. For magnetic separators with measured mass flow rates described by $\varepsilon(d)$, cross-validation yields a negative R_{LOO}^2 , supporting the use of $\eta(Fe)$ in (3). Leave-one-sample-out cross-validation is a logical next step in extending the predictive-performance evaluation of $\eta(Fe)$ for magnetic separators.

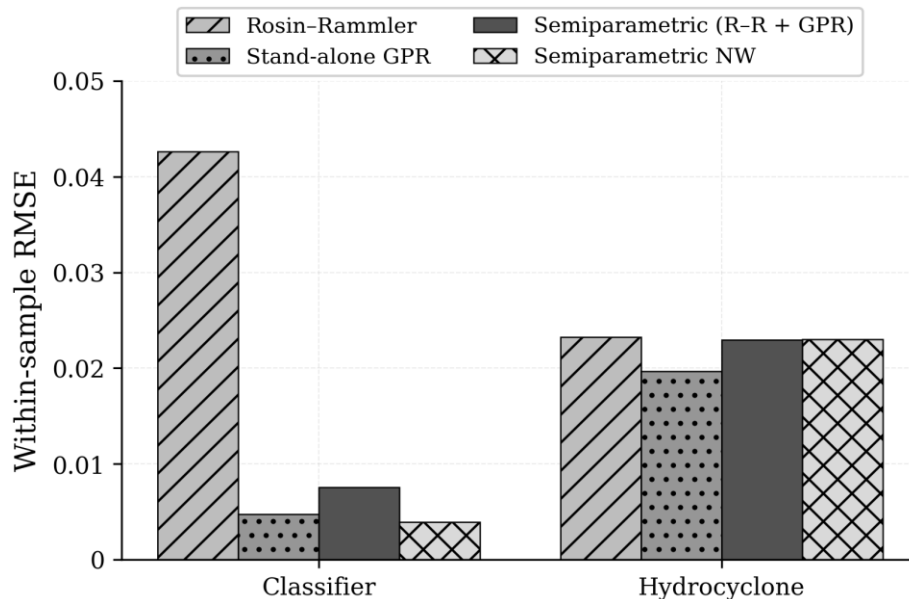


Fig. 2. Comparison of within-sample model RMSE for size-separation units with an identified mass-flow-rate ratio

The robustness of identification to measurement errors is evaluated by Monte Carlo simulation. A total of $N = 500$ replicates (seed = 42) are generated with additive noise under the specified model, whose standard deviation is $\sigma = \max(0.02, 0.05 \varepsilon)$. The complete identification procedure is applied to every replicate (Fig. 3). The resulting bands, defined by the 2.5th and 97.5th percentiles across replicates, characterize the sensitivity of the identified characteristic to the adopted noise model. For the classifier, the mean relative band width is approximately 14 % of the mean curve value; it increases to approximately 21 % for units with fewer data points. The median RMSE between a replicate and the baseline curve does not exceed 0.03.

Thus, cross-validation characterizes the model's ability to generalize, while Monte Carlo simulation quantifies the sensitivity of the identified characteristic to the adopted measurement-error model. Together, these results characterize the predictive performance and robustness of the proposed method.

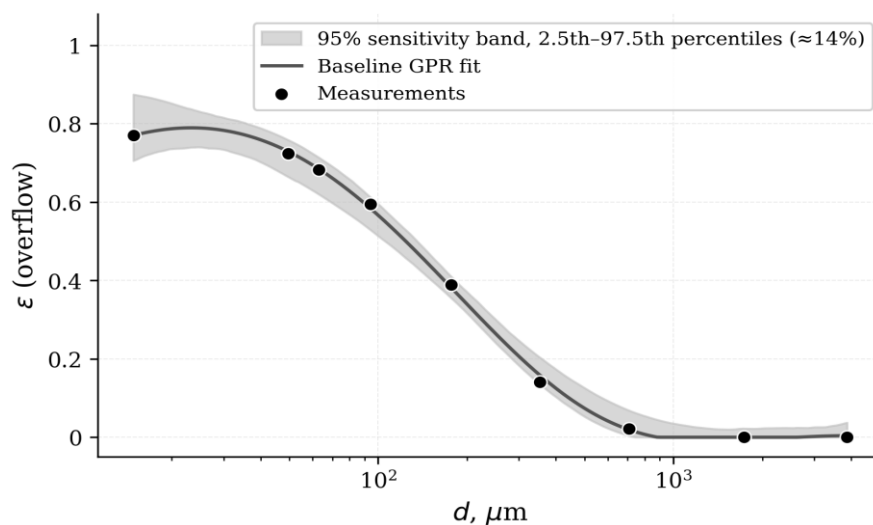


Fig. 3. Monte Carlo simulation

Conclusions

A semiparametric method has been developed for identifying the separation characteristics of iron ore beneficiation processes. The method combines a physically motivated prior mean—the decreasing Rosin–Rammler model with the $\ln 2$ factor for size-class recovery to the overflow, the increasing sigmoid model for magnetic separation

as a function of iron content, or an empirical-mean baseline for units operating in the saturation region—with a nonparametric correction term estimated by GPR from empirical deviations of observations from this function. In the Rosin–Rammler model, d_{50} is the median cut size. Depending on the amount of informative data, the estimate interpolates between the prior mean and the correction without discrete switching between methods. The same estimator is used for full-sample estimation, cross-validation, and Monte Carlo analysis; a semiparametric NW estimator constructed from the deviations r_i serves as a benchmark. Experimental recovery is defined consistently for all models using the same mass-balance equation.

The provenance of the mass-flow-rate ratios r_Q used in the mass-balance definition of ε is documented. The principal quantitative result is obtained for the classifier, for which $r_Q = 2.08$ is well constrained by the measured feed data, with a 95 % bootstrap interval of [1.95, 2.42]. The proposed GPR-based semiparametric model reduces within-sample RMSE by approximately 82 % relative to the Rosin–Rammler model and shows better predictive performance than the stand-alone GPR model under leave-one-out cross-validation: $R_{LOO}^2 = 0.970$ versus 0.945. The ratio r_Q is re-estimated in every cross-validation iteration without using data from the held-out size class.

For the hydrocyclone, r_Q is estimated with a 95 % bootstrap interval of [0.96, 2.56]. The proposed model yields the highest $R_{LOO}^2 \approx 0.81$ among the compared models throughout this range of r_Q , while the corresponding d_{50} range of approximately 16–30 μm quantifies the sensitivity of this parameter estimate to the mass-flow-rate ratio. Results for magnetic separators with measured mass flow rates confirm the importance of selecting the argument of the separation characteristic according to the physical separation mechanism: $R_{LOO}^2 < 0$ for $\varepsilon(d)$ supports the use of $\eta(\text{Fe})$. All reported numerical values can be reproduced using a single script with a fixed pseudorandom-number generator state and are stored in machine-readable artifacts containing the metrics, curves, flow-data provenance, and sensitivity analysis.

The method provides the model and uncertainty information required for model predictive control: point estimates and uncertainty measures can be supplied to local MPC optimization problems and to a coordinator that exchanges uncertainty estimates among agents. Further development includes cross-validation of $\eta(\text{Fe})$, closed-loop MPC experiments to assess the effect of uncertainty on control guarantees, and online adaptation of the identified characteristics when feed mineralogy and processing properties change.

ADDITIONAL INFORMATION

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, formulation of research objectives, development of the semiparametric identification method and models, software, validation, analysis of results, visualization, writing – original draft preparation, writing – review and editing – V.T.

DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS

In preparing this work, the author used OpenAI Codex in a limited manner for language polishing, sentence-level translation, terminology harmonization, and formatting assistance. The use of this tool was strictly limited to improving grammar, clarity, and stylistic consistency. The author independently developed the scientific ideas, methodology, mathematical models, algorithms, software, experimental design, interpretation of results, and conclusions. The final manuscript was critically reviewed and validated by the author, who remains fully accountable for its content.

REFERENCES

1. Wills B. A., Finch J. A. Wills' Mineral Processing Technology: An Introduction to the Practical Aspects of Ore Treatment and Mineral Recovery. 8th ed. Oxford: Butterworth-Heinemann, 2015. 512 p. <https://doi.org/10.1016/C2010-0-65478-2>
2. Botha S., le Roux J. D., Craig I. K. Hybrid non-linear model predictive control of a run-of-mine ore grinding mill circuit. *Minerals Engineering*. 2018. Vol. 123. P. 49–62. <https://doi.org/10.1016/j.mineng.2018.04.016>
3. Tambun R., Furukawa K., Hirayama M., Shimadzu M., Yamanaka S., Ohira Y. Measurement and Estimation of the Particle Size Distribution by the Buoyancy Weighing–Bar Method and the Rosin–Rammler Equation. *Journal of Chemical Engineering of Japan*. 2016. Vol. 49, No. 2. P. 229–233. <https://doi.org/10.1252/jcej.14we129>
4. McCoy J. T., Auret L. Machine learning applications in minerals processing: A review. *Minerals Engineering*. 2019. Vol. 132. P. 95–109. <https://doi.org/10.1016/j.mineng.2018.12.004>
5. Rasmussen C. E., Williams C. K. I. Gaussian Processes for Machine Learning. Cambridge, MA: The MIT Press, 2006. 272 p.
6. Härdle W. Applied Nonparametric Regression. Cambridge: Cambridge University Press, 1990. 333 p.
7. Wasserman L. All of Nonparametric Statistics. New York: Springer, 2006. 268 p.

8. Silverman B. W. Density Estimation for Statistics and Data Analysis. London: Chapman and Hall, 1986. 175 p.

9. Morkun V. S., Morkun N. V., Tron V. V. Identification of control systems for ore-processing industry aggregates based on nonparametric kernel estimators. *Metallurgical and Mining Industry*. 2015. Vol. 7, No. 1. P. 14–17.

10. Tron V. V. Analiz metodiv identyfikatsii system avtomatyzovanoho keruvannia tekhnolohichnykh protsesamy zbahachennia zalizorudnoi syrovyny [Analysis of identification methods for automated control systems of iron ore beneficiation processes]. *Visnyk Kryvorizkoho natsionalnoho universytetu*. 2013. Issue 35. P. 198–201. (in Ukrainian).

11. Hewing L., Kabzan J., Zeilinger M. N. Cautious model predictive control using Gaussian

process regression. *IEEE Transactions on Control Systems Technology*. 2020. Vol. 28, No. 6. P. 2736–2743. <https://doi.org/10.1109/TCST.2019.2949757>

12. Efron B., Tibshirani R. J. An Introduction to the Bootstrap. New York: Chapman & Hall, 1993. 436 p.

13. Newey W. K., West K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*. 1987. Vol. 55, No. 3. P. 703–708. <https://doi.org/10.2307/1913610>

14. Politis D. N., Romano J. P. The stationary bootstrap. *Journal of the American Statistical Association*. 1994. Vol. 89, No. 428. P. 1303–1313. <https://doi.org/10.1080/01621459.1994.10476870>

Віталій ТРОНЬ

Криворізький національний університет

НАПІВПАРАМЕТРИЧНА ІДЕНТИФІКАЦІЯ ТА КІЛЬКІСНЕ ОЦІНЮВАННЯ НЕВИЗНАЧЕНОСТІ СЕПАРАЦІЙНИХ ХАРАКТЕРИСТИК ЗБАГАЧУВАЛЬНИХ АПАРАТІВ

Ідентифікація сепараційних характеристик апаратів збагачення залізної руди за виробничими даними формує інформацію про модель і невизначеність, необхідну для модельно-прогнозного керування. Параметричні моделі, зокрема модель Розіна–Рамлера та сигмоїдна модель, мають фіксовані аналітичні форми, які можуть недостатньо точно описувати характеристики окремих апаратів. Точкова ідентифікація без окремої статистичної процедури не дає кількісної оцінки невизначеності прогнозу, тоді як суто непараметричні оцінки можуть бути нестійкими за обмежених даних. У роботі розроблено уніфікований метод напівпараметричної ідентифікації, у якому фізично обґрунтовану параметричну функцію апіорного середнього доповнено коригувальною складовою, оціненою регресією гауссових процесів за емпіричними відхиленнями спостережуваних значень. Для оцінювання моделі, крос-валідації та аналізу чутливості застосовано єдину процедуру. Напівпараметричний оцінювач Надарай–Уотсона, побудований за відхиленнями від функції апіорного середнього, використано як порівняльну модель. Експериментальне вилучення обчислено за єдиним масово-балансовим означенням. Співвідношення масових витрат продуктів визначено за результатами вимірювань або оцінено за покласовим балансом. Для класифікатора це співвідношення надійно визначено за даними, а його 95% бутстреп-інтервал становить [1,95; 2,42]. Запропонована напівпараметрична модель знижує внутрішньовибіркову середньоквадратичну похибку приблизно на 82% порівняно з моделлю Розіна–Рамлера та перевершує окрему модель регресії гауссових процесів за результатами крос-валідації з виключенням одного класу крупності: коефіцієнти детермінації становлять відповідно 0,970 і 0,945. Співвідношення масових витрат переоцінювали на кожній ітерації крос-валідації з використанням лише навчальних класів крупності. Для гідроциклона 95% бутстреп-інтервал цього співвідношення становить [0,96; 2,56], а запропонована модель зберігає найвищий коефіцієнт детермінації за крос-валідацією — близько 0,81 — у всьому інтервалі. Результати для магнітних сепараторів із вимірними масовими витратами підтверджують, що аргумент сепараційної характеристики має відповідати фізичному механізму розділення. Від’ємні коефіцієнти детермінації, отримані під час моделювання сепараційної характеристики як функції крупності частинок, свідчать про непридатність крупності як її аргументу та обґрунтовують подання цієї характеристики як функції вмісту заліза, що безпосередніше відображає механізм магнітного розділення. Моделювання Монте-Карло за заданою моделлю шуму застосовано для кількісного оцінювання чутливості; для класифікатора середня відносна ширина смуги невизначеності становить приблизно 14%.

Ключові слова: напівпараметрична ідентифікація; сепараційна характеристика; регресія гауссових процесів; ядерна корекція; матеріальний баланс; крос-валідація; оцінювання невизначеності; відтворюваність; збагачення залізної руди.