Taylor & Francis
Taylor & Francis Group

Science and Technology of Advanced Materials: Methods

ISSN: 2766-0400 (Online) Journal homepage: www.tandfonline.com/journals/tstm20

Replica-exchange Bayesian mixture regression reveals cluster-dependent XRD descriptors of tensile modulus in recycled polypropylene

Kazuki Hammura, Kiyotaka Hitomi, Kenji Nagata & Masanobu Naito

To cite this article: Kazuki Hammura, Kiyotaka Hitomi, Kenji Nagata & Masanobu Naito (2026) Replica-exchange Bayesian mixture regression reveals cluster-dependent XRD descriptors of tensile modulus in recycled polypropylene, Science and Technology of Advanced Materials: Methods, 6:1, 2662046, DOI: 10.1080/27660400.2026.2662046

To link to this article: https://doi.org/10.1080/27660400.2026.2662046

SCIENCE AND TECHNOLOGY OF ADVANCED MATERIALS: METHODS
2026, VOL. 6, NO. 1, 2662046
https://doi.org/10.1080/27660400.2026.2662046

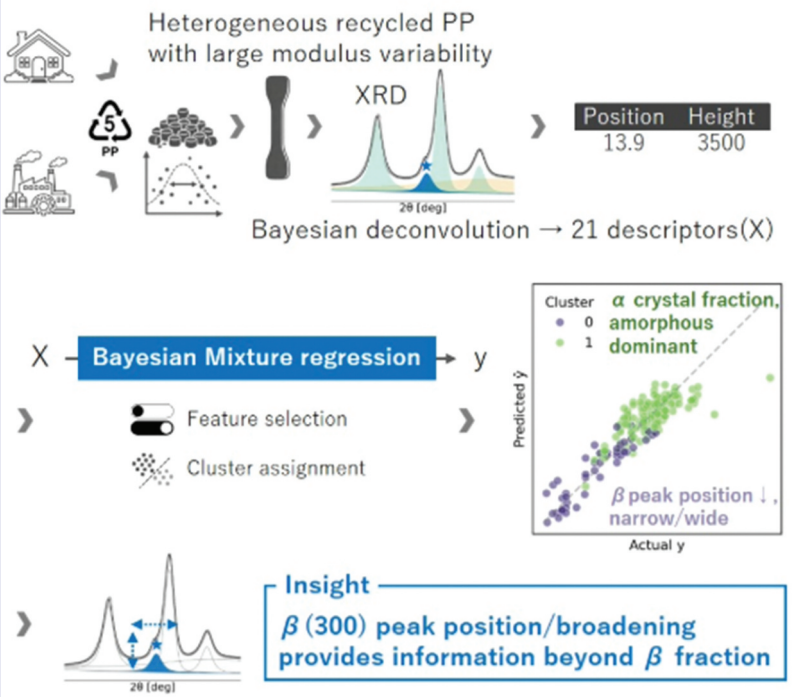
Replica-exchange Bayesian mixture regression reveals cluster-dependent XRD descriptors of tensile modulus in recycled polypropylene

Kazuki Hammura [a,b], Kiyotaka Hitomi [c], Kenji Nagata [d] and Masanobu Naito [a,c]

[a]Graduate School of Science and Technology, University of Tsukuba, Ibaraki, Japan; [b]Platform Laboratory for Science and Technology, Asahi-Kasei Corporation, Tokyo, Japan; [c]Research Center for Macromolecules and Biomaterials, National Institute for Materials Science (NIMS), Ibaraki, Japan; [d]Center for Basic Research on Materials, National Institute for Materials Science (NIMS), Ibaraki, Japan

ABSTRACT

Recycled polypropylene (rPP) exhibits large property variability due to mixed origins and degradation histories, complicating nondestructive grading. In this study, we propose an interpretable Bayesian framework that links X-ray diffraction (XRD) peak features to tensile modulus for virgin/recycled PP blends subjected to xenon-arc weathering. XRD profiles were analyzed by Bayesian peak deconvolution, extracting physically interpretable descriptors from four low-angle crystalline peaks (α(110), α(040), α(130), β(300)) and a broad amorphous halo, yielding 21 explanatory variables per sample. A Bayesian finite mixture of linear regressions with probabilistic feature selection was fitted, and posterior inference using replica-exchange Monte Carlo was performed to explore a highly multimodal posterior. The model selected two clusters and achieved an in-sample fit ($R^2 = 0.81$, RMSE = 145 MPa). Replicate-holdout group k-fold cross-validation provided a conservative generalization estimate at the tensile level ($R^2 = 0.15$, RMSE = 320 MPa, $N = 120$), providing a conservative lower-bound estimate due to specimen mismatch and repeated labels at 0 cycles. Clusters differed in the β(300) descriptor space, and direct comparison of cluster-specific posterior coefficient distributions indicated that the β(300) peak position provided the clearest evidence of cluster-dependent regression behavior, whereas peak broadening was relevant in both clusters. These results suggest that β(300)-related descriptors – potentially reflecting β-phase lattice strain or local disorder – may contribute to modulus beyond β fraction alone. This framework provides interpretable XRD descriptors and uncertainty-aware modulus estimates for grading heterogeneous rPP.



IMPACT STATEMENT

Combining Bayesian XRD peak deconvolution with replica-exchange Bayesian mixture regression and probabilistic feature selection, we enable uncertainty-aware, cluster-dependent grading of recycled polypropylene and identify β(300) peak-state descriptors of tensile modulus.

ARTICLE HISTORY

Received 24 January 2026
Revised 20 March 2026
Accepted 15 April 2026

KEYWORDS

Recycled polypropylene; X-ray diffraction; Bayesian mixture regression; Bayesian peak deconvolution; structure–property relationship

CONTACT Masanobu Naito NAITO.Masanobu@nims.go.jp Research Center for Macromolecules and Biomaterials, National Institute for Materials Science (NIMS), 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

Supplemental data for this article can be accessed online at https://doi.org/10.1080/27660400.2026.2662046

1. Introduction

The transition from a linear to a circular economy is one of the most pressing global challenges. To achieve this goal, polymer recycling is a key technology. Currently, there are two primary approaches to polymer recycling: material (mechanical) and chemical (depolymerization) recycling. The former is generally preferred because it typically requires less energy and fewer resources than chemical depolymerization. However, a pivotal challenge is the large variability of recycled polymers arising from mixed origins, additives, contamination, and different degradation histories [1–4].

Polypropylene (PP) is a light and durable plastic widely used in various applications such as packaging, automotive parts, and household goods. The global PP production was 81.9 million tons in 2024, the highest among commodity plastics [5]. Even for virgin PP (vPP), quality control is nontrivial because its crystalline and mechanical properties are governed by multiple, partially redundant structural factors, including crystallinity, polymorphism (α/β), and crystallite size [6,7]. Specifically, the polymorphic composition plays a crucial role; it is generally established that polypropylene dominated by the β-crystalline phase exhibits a lower tensile modulus compared to its α-dominated counterpart. While β-PP may show higher elongation and tensile strength, literature comparisons indicate that its elastic modulus and yield stress are typically lower than those of α-PP [6].

For recycled PP (rPP), this challenge of structural variability is substantially amplified. In addition to the intrinsic complexity arising from crystalline polymorphism mentioned above, rPP exhibits much larger variability due to differences in feedstock origin, processing history, additives, and degradation state. In particular, post-consumer recycled (PCR) PP typically originates from heterogeneous waste streams and therefore shows large fluctuations in molecular weight, contamination level, and crystalline structure, whereas post-industrial recycled (PIR) PP is generally more homogeneous but can still diverge from vPP because of repeated thermal and mechanical histories. Several studies have reported that rPP differs from vPP not only at the molecular level but also in higher-order and mesoscale internal structures [8,9].

Moreover, aging and photo-oxidation can induce molecular chain scission and subsequent structural reorganization, further modifying crystalline states and mechanical response [10,11]. As a result, similar physical properties such as tensile modulus of PP can emerge from distinct combinations of structural variables, underscoring the need for analysis frameworks that go beyond single scalar descriptors.

Tensile modulus is a key property for PP, but it is typically obtained by destructive testing and is therefore unavailable for in-service products. Thus, nondestructive measurements that enable property prediction are of practical interest for screening and grading diverse rPP materials. X-ray diffraction (XRD) directly probes higher-order structure (crystallinity, polymorphs, and lattice disorder) that governs stiffness, while vibrational spectroscopic methods (e.g. near-infrared spectroscopy (NIR)) primarily reflect chemical and chain-level information [12]. In this context, XRD profiles contain richer structural information than NIR spectra. However, raw XRD profiles are high-dimensional, and it is well recognized that peak-deconvoluted parameters provide more insight into physical properties compared to the raw profiles. Although peak deconvolution is often performed manually and suffers from low throughput, the physically meaningful factors obtained by peak deconvolution are entangled and often correlated (e.g. crystallinity, α/β polymorph contributions, and peak broadening). This motivates a modeling approach that extracts a compact set of interpretable XRD descriptors and allows for cluster-dependent structure-modulus mappings, rather than relying solely on a single global predictive relationship.

Regression and classification of material properties from profile- or spectrum-type data are central tasks in materials informatics, in addition to process optimization, in-silico design, target screening, and design of experiments [13]. Many workflows rely on off-the-shelf ('toolbox') models; however, using an entire spectrum/profile as input is challenging because the variables are high-dimensional and strongly collinear. Conventional approaches such as partial least squares regression and support vector machines can be effective in some settings but may become less robust when the dataset spans multiple domains or exhibits regime shifts [14,15]. In polymer materials, interpretability is particularly important for linking model outputs to the hierarchical structure of matter and for extracting a compact set of actionable descriptors from limited data [13].

Bayesian modeling provides a reasonable route to address these constraints by collectively handling uncertainty, multicollinearity, and model complexity also within materials science [16,17]. In particular, a mixture regression equipped with binary feature-gating variables enables probabilistic feature selection and interpretability via posterior inclusion probabilities (PIPs) of physically meaningful XRD descriptors. In our current implementation, the feature-gating prior is uniform over inclusion/exclusion, so feature usage is learned from the data rather than enforced by a strong sparsity-inducing prior. Additionally, a mixture model can address nonlinear transitions between clusters and capture cluster-dependent contributions. Bayesian approaches are well-suited for

inferring latent spectral shapes [18,19]. Furthermore, the use of replica-exchange Monte Carlo (REMC) enables an effective exploration of multimodal parameter spaces [20]. By preventing the sampling process from being trapped in local optima, this framework enables more robust exploration of the posterior distribution.

In this work, we propose a Bayesian framework that links physically interpretable XRD peak features derived from Bayesian peak deconvolution to the tensile modulus for PP specimens with different virgin/recycled compositions and xenon-arc weathering cycles. The model is a Bayesian finite mixture of linear regressions with binary feature selection, and posterior inference is performed using REMC to explore multimodal parameter spaces.

Our contributions are:

- Method: development of an interpretable Bayesian mixture regression with probabilistic feature selection for analytical descriptors from small datasets.
- Application: the derivation of interpretable higher-order structure-modulus mapping from various compositions and aging states of PP test pieces.
- Insight: identification of two structural clusters with different contributions of crystalline/amorphous descriptors (including β crystal-related features), providing a mechanistic view of variability in rPP.

2. Materials and methods

2.1. Materials and experimental design

Two types of vPP (Prime Polypro® E100-GPL from Prime Polymer Co., Ltd. and Novatec® EA9 from Japan Polypropylene Corporation) and one rPP (PP RP N, Nakasada jushi) were purchased. The Nakasada product was recycled from factory recovered medical plastics. Preparation of dumbbell-shaped test pieces (JIS K7139 A12) [21] was carried out using an Xplore twin screw compounder, extruder, and injection molding machine. The composition ratios of vPP and rPP were set to 75:25, 50:50, 25:75, and 0:100. In addition to this sample group, another rPP material (hard PP, Toyama Kankyo Seibi Co., Ltd.) was also molded into the same size test piece. Toyama rPP was used as a single material for dumbbell test piece. This material was derived from food packaging waste (random PP) and contained small amounts of polyethylene (PE). Nakasada and Toyama are classified as PIR and PCR, respectively. Table 1 shows the composite patterns. Fifteen test pieces were prepared for each composition. Twelve specimens per composition were assigned to the following aging treatment. The remaining three specimens were kept unaged (0 cycles; $n = 3$) and used for tensile testing only. Aging treatment with heat, light, and mist was performed with a xenon-arc weathering tester following JIS K7350-2 cycle No.10 (1 cycle = 2 h) [22] for 30, 50, 70, and 100 cycles.

2.2. XRD measurement

For these specimens, XRD profiles were measured both before and after the aging treatment. XRD measurements were performed using a MiniFlex (Rigaku Corporation, Japan) with Cu Kα radiation operated at 40 kV and 15 mA. Diffraction patterns were collected in Bragg-Brentano (θ-2θ) reflection geometry over $2\theta = 7.5–32.5°$ with a step size of 0.01° and a scan speed of 0.5°/min. Full settings are provided in Table S1. In total, 192 XRD profiles were acquired (8 compositions × 4 aging levels × 3 replicates × 2 time points (before and after aging)). Representative XRD profiles showing the variations in peak intensity and amorphous halo across different compositions and aging conditions are shown in Figure 1.

2.3. Tensile modulus measurement

Tensile modulus was measured using AG-X Plus (Shimadzu Corporation, Japan) according to JIS K 7162. Tests were conducted at a crosshead speed of 50 mm/min. The modulus was determined from the maximum slope in the initial linear region of the stress–strain curve using a two-point method, as provided by the testing software. Because tensile testing is destructive, tensile modulus and XRD were not always measured on the same physical specimen. For the weathered conditions (30, 50, 70, and 100 cycles), XRD profiles after aging were measured on the same specimens prior to tensile testing. For the unaged condition (0 cycles), tensile modulus was measured on the three unaged specimens for each composition without XRD, and these values were paired with the pre-weathering XRD profiles (measured before exposure) by composition and nominal replicate index (1–3). Here, the replicate index was used only as an identifier to associate the 0-cycle tensile measurements with the pre-weathering XRD profiles of specimens assigned to 30, 50, 70, and 100 cycles; thus, XRD and tensile were not measured on the same physical specimen at 0 cycles. Consequently, for each composition-replicate combination, four pre-weathering XRD-derived feature vectors were associated with a single 0-cycle tensile modulus value. In total, 192 XRD profiles and 120 tensile modulus measurements were used for regression.

2.3.1. Overview of the analysis workflow

Figure 2 summarizes the overall analysis workflow used in this study. The objective of the workflow is to identify XRD-derived structural descriptors that explain the tensile modulus of recycled polypropylene and its composites.

The analysis consists of two stages. In the first stage, each measured XRD profile is analyzed using Bayesian peak deconvolution. This procedure decomposes the diffraction profile into crystalline peaks, an amorphous halo, and background, and yields posterior estimates of peak parameters such as peak position, width, area, and height. For each XRD profile, one 21-dimensional descriptor vector was extracted from the lowest-energy posterior sample in the coldest chain ($\beta = 1$). For the unaged condition, the descriptor vectors used in regression were those extracted from the pre-weathering XRD profiles and paired with the 0-cycle tensile measurements as described in Overview of the analysis workflow.

In the second stage, the extracted XRD descriptors are related to the experimentally measured tensile modulus using a Bayesian mixture regression model with built-in feature selection. The response variable in this regression is the experimentally measured tensile modulus itself; it is not inferred in the deconvolution stage. Thus, Bayesian peak deconvolution is used solely as a feature-extraction step, and the current implementation does not propagate the full deconvolution posterior into the regression stage.

This two-stage framework enables physically interpretable structural descriptors derived from XRD data to be linked with mechanical properties through a probabilistic regression model.

2.5. Bayesian peak deconvolution

Following previous work [23,24], Bayesian peak deconvolution was applied to each diffraction profile to extract physically interpretable peak descriptors. The

Table 1. Composition of PP dumbbell test specimens.

	Virgin PP		Recycled PP		
Group No.	Kind	Ratio	Kind	Ratio	Sum
1	EA9	25	Nakasada	75	100
2		50		50	100
3		75		25	100
4	E100-GPL	25		75	100
5		50		50	100
6		75		25	100
7	–	0		100	100
8	–	0	Toyama	100	100

Compositions of PP dumbbell specimens used in this study. Each formulation was molded into fifteen replicate JIS K7139 A12 test pieces: $n = 3$ for each of five aging levels (0, 30, 50, 70, and 100 cycles). One recycled PP (Nakasada) was blended with virgin PP at different mixing ratios, whereas the other recycled PP (Toyama) was molded as a single-material specimen.

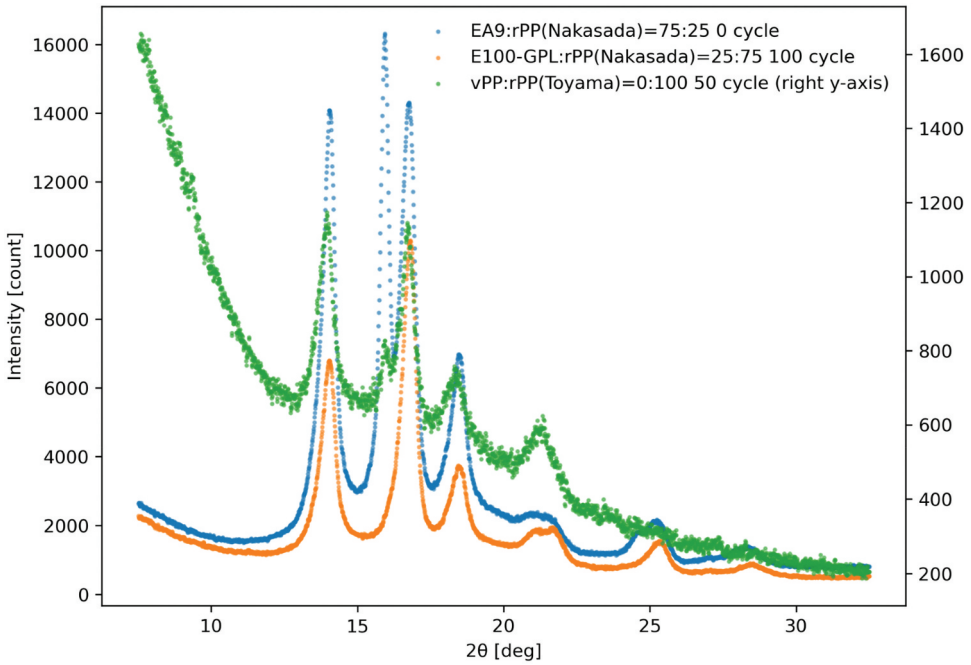


Figure 1. Representative XRD profiles from the dataset. Three representative XRD profiles are shown to illustrate the diversity of structural states in the dataset, including variations in crystalline peak intensity and amorphous halo contribution across polymer compositions and weathering conditions. The Toyama rPP profile (green) is plotted against the right y-axis for visibility.

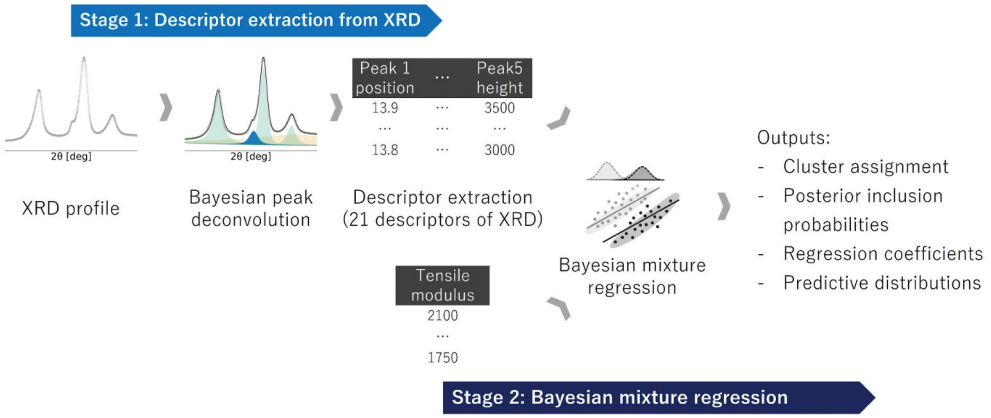


Figure 2. Overview of the two-stage analysis workflow. In Stage 1, each XRD profile was analyzed using Bayesian peak deconvolution to extract physically interpretable peak descriptors. In Stage 2, one 21-dimensional descriptor vector per profile was used as the explanatory input to a Bayesian mixture regression model with cluster-specific probabilistic feature selection, while the response variable was the experimentally measured tensile modulus. The model jointly inferred cluster assignments, posterior inclusion probabilities, and cluster-specific regression coefficients.

deconvolution with peak selection was performed using a common full set of initial peak-center values determined from all observed profiles. All profiles were successfully deconvoluted, including severely aged samples that exhibited a pronounced amorphous halo and weak crystalline peak intensities. A representative result of the Bayesian peak deconvolution is shown in Figure 3. The observed profiles were well reconstructed by the sum of crystalline peaks, an amorphous halo, and background, and the adequacy of the fit was confirmed by inspection of the residuals between the observed and reconstructed profiles. The calculations were performed using a series of Python scripts ported from the original C code provided by Nagata.

Four major low-angle crystal peaks (α(110), α(040), α(130), and β(300)) and a broad amorphous halo centered at ~17.5° in 2θ were selected as sources for feature extraction. For each of these five components, peak center position, relative area, height, and full width at half maximum (FWHM) were calculated. In addition, overall crystallinity and β crystallinity were calculated from the relative peak areas, and β height fraction was calculated from the peak heights, following a previous report [25].

For downstream regression analysis, each XRD profile was represented by a single descriptor vector defined by the deconvolution parameter set of the lowest-energy posterior sample in the coldest chain ($\beta = 1$). Thus, in the current implementation, Bayesian peak deconvolution serves solely as a feature-extraction step: tensile modulus is not inferred in this stage, and the full deconvolution posterior distribution is not propagated into the subsequent mixture regression. After excluding variables that were multicollinear by definition, the final descriptor set comprised 21 variables per sample (5 peak positions (2θ), 3 relative areas, 5 FWHMs, 5 peak heights, and 3 crystallinity-related quantities; Table S2). These descriptors were used as explanatory variables in the Bayesian mixture regression model described in the following section, while the response variable was the experimentally measured tensile modulus.

All raw data were stored in Research Data Express (RDE; https://dice.nims.go.jp/services/RDE/) at the National Institute for Materials Science, which supports flexible data sharing and reuse.

2.3.2. Bayesian mixture regression modeling

A finite mixture of linear regression models with built-in feature selection was designed and trained using Bayesian inference. The model is designed to address two common challenges in XRD-based structure–property regression: (1) multicollinearity among analytically derived, physically interpretable peak descriptors, and (2) latent heterogeneity in the structure–property relationship that motivates cluster-dependent regression mappings. Specifically, the model combines probabilistic feature gating with latent cluster assignment of samples to mixture components. A graphical representation of the model is shown in Figure 4. Throughout Sections 2.6-2.9, X_i denotes the D-dimensional XRD descriptor vector for sample i, X_{id} its D-th component, $c_k = (c_{k1}, \ldots, c_{kD})^\top$ the cluster-specific feature-selection vector, and $w_k = (w_{k1}, \ldots, w_{kD})^\top$ the cluster-specific regression-coefficient vector. This figure summarizes the generative process; we now define the likelihood and priors corresponding to each node and describe the posterior inference procedure. Posterior inference was performed using REMC, which enables efficient exploration of a broad parameter space. This approach was adopted because the simultaneous inference of

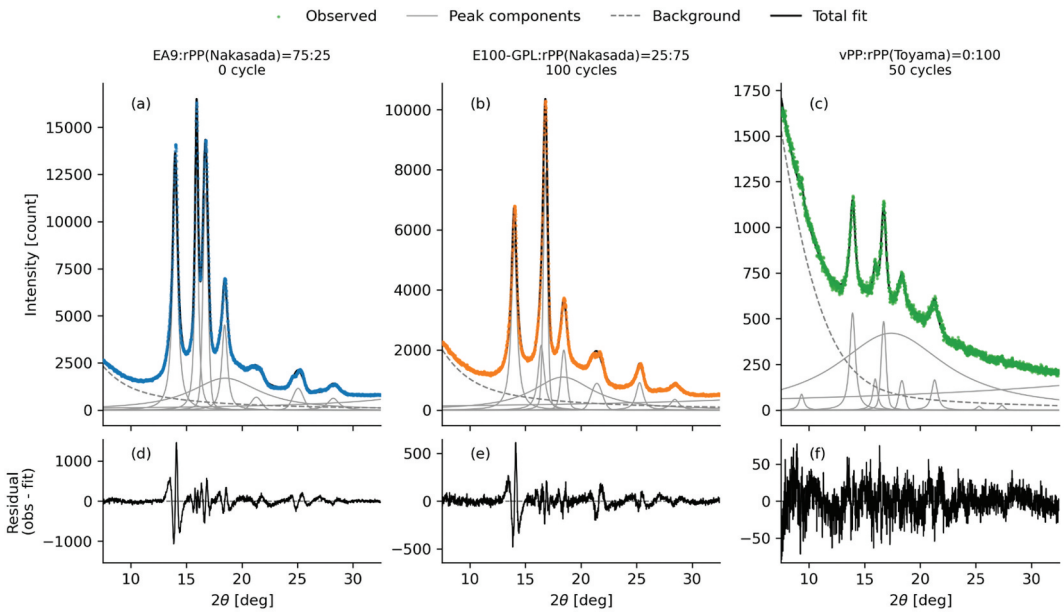


Figure 3. Representative Bayesian XRD peak-deconvolution results and corresponding residuals. The same three representative XRD profiles as in Figure 1 are shown. In panels (a-c), observed diffraction profiles are shown as symbols, together with the individual fitted components including crystalline peaks and the amorphous halo (thin gray lines), the background (dashed line), and the total fitted profile (solid line). Panels (d-f) show the residuals (observed – fitted) for the corresponding profiles.

cluster assignments and feature selection leads to a highly multimodal posterior distribution, for which conventional single-chain sampling methods are often insufficient.

2.3.3. Model specification

Let $i = 1, \ldots, N$ index samples, $k = 1, \ldots, K$ index clusters, and $d = 1, \ldots, D$ index XRD features. Let $X_i = (X_{i1}, \ldots, X_{iD})^\top$ denote the XRD feature vector for sample i. Mixture weights are drawn as $\pi \sim \text{Dirichlet}(\alpha)$, and cluster assignments follow $z_i \mid \pi \sim \text{Categorical}(\pi)$. For each cluster k, feature-selection indicators satisfy $c_{kd} \sim \text{Bernoulli}(0.5)$, regression weights satisfy $w_{kd} \sim \mathcal{N}(0, \sigma_w^2)$, and the intercept satisfies $b_k \sim \mathcal{N}(0, \sigma_b^2)$. Define the cluster-specific vectors $c_k = (c_{k1}, \ldots, c_{kD})^\top$, $w_k = (w_{k1}, \ldots, w_{kD})^\top$, and $\theta_k = c_k \odot w_k$. Given $z_i = k$, the tensile modulus is modeled as $y_i \mid z_i = k, X_i, c_k, w_k, b_k \sim \mathcal{N}(X_i^\top \theta_k + b_k, \sigma_R^2)$, where $\odot$ denotes the elementwise product, so that $\theta_{kd} = c_{kd} w_{kd}$. In this study, we fix the inclusion prior to be symmetric, $p(c_{kd} = 1) = p(c_{kd} = 0) = 0.5$, i.e. a uniform prior over inclusion/exclusion without an explicit sparsity-inducing penalty.

2.3.4. Posterior inference via REMC

Posterior inference was performed using REMC with likelihood tempering over inverse temperatures $\beta \in [0, 1]$. We used 24 chains with $\beta_0 = 0$ and $\beta_{23} = 1$, and intermediate β values were generated by a geometric schedule. Neighboring chains were proposed to swap every iteration, with acceptance based on the per-sample negative log-likelihood E (energy) as implemented in the code. Within each chain, we alternated updates of (i) mixture weights by a Dirichlet draw using tempered counts $\alpha + \beta n_k$, (ii) feature indicators c_{kd} by single-bit flip Metropolis steps, (iii) regression coefficients and intercepts by random-walk Metropolis steps, and (iv) latent assignments by softmax sampling. All inference hyperparameters are summarized in Table 2.

To address the label switching problem inherent in finite mixture models, we applied a post-processing step to ensure cluster identifiability. For each posterior sample, cluster labels were reordered based on the mean predicted tensile modulus of the samples assigned to each cluster (ascending order). Consequently, 'Cluster 0' consistently refers to the regime with lower modulus, and 'Cluster 1' to the higher modulus regime.

2.3.5. Out-of-sample prediction

Because cluster assignments in the mixture model are inferred conditional on the observed tensile modulus, we evaluate out-of-sample prediction without using test labels by marginalizing the latent cluster. Specifically, for a new feature vector X_*, the predictive mean is computed as $\hat{y}(X_*) = \sum_{k=1}^{K} \pi_k \mu_k(X_*)$, where $\mu_k(X_*) = X_*^\top \theta_k + b_k, \theta_k = c_k \odot w_k$. Predictive intervals are obtained from the posterior predictive distribution.

3. Results

3.1. Dataset overview

Test pieces prepared from the Toyama rPP (PCR) exhibited lower tensile modulus values compared with those prepared from the Nakasada rPP and its

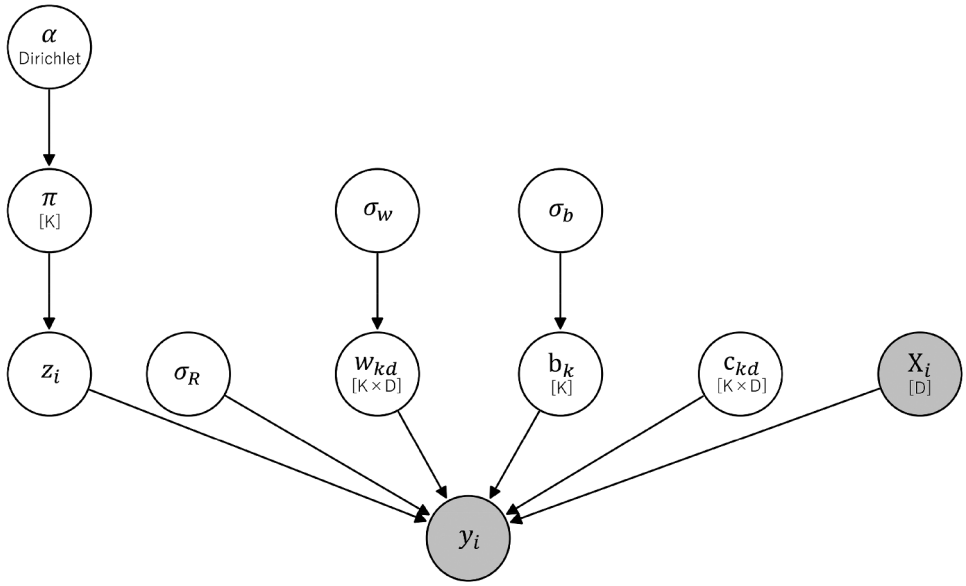


Figure 4. Graphical model of the Bayesian finite mixture of linear regressions with cluster-specific feature selection. Each sample i is assigned to a latent cluster z_i according to mixture weights π drawn from a Dirichlet prior. Conditional on $z_i = k$, the tensile modulus y_i is modeled as a linear function of the XRD-derived feature vector X_i, with cluster-specific regression weights w_{kd}, intercept b_k, and binary inclusion indicators c_{kd} for descriptor d. Thus, the set of active descriptors can differ across clusters. Here, $d = 1, \ldots, D$ indexes XRD descriptors. σ_R denotes the observation-noise standard deviation, and σ_w and σ_b denote the prior standard deviations for regression weights and intercepts, respectively.

Table 2. Model parameter settings.

Category	Setting	Used value	Notes
Mixture components	K	2	Selected by free energy comparison across K (Fig. S1).
Dirichlet prior	α	1	$\pi \sim Dirichlet(\alpha)$
Noise (scaled space)	σ_R	0.35	Standard deviation in standardized y space.
Coef prior (scaled)	σ_w	0.4	$w_{kd} \sim \mathcal{N}(0, \sigma_w^2)$
Intercept prior (scaled)	σ_b	0.7	$b_k \sim \mathcal{N}(0, \sigma_b^2)$
Temperatures	β ladder	24 chains, $\beta_0 = 0$, $\beta_{last} = 1$	Geometric schedule with proportion 1.10
Burn-in	Iterations	10,000	Per chain (burn-in exchanges = burn_in/exchange_frequency).
Sampling	Iterations	10,000	Per chain after burn-in.
Exchange	Frequency	Every 1 iteration	Adjacent swaps, alternating pairs.

The same inference settings were used per fold; standardization was fit on training data only.

composites with vPPs. From a polymorph perspective, in the samples before aging, a smaller amount of PP β-crystal was observed in the Nakasada materials than vPPs, whereas a few PE crystalline peaks appeared only in the Toyama rPP.

Figure 1 shows representative XRD profiles from the dataset. These diffraction patterns illustrate the characteristic crystalline peaks of polypropylene together with the amorphous halo associated with disordered regions. Differences in peak intensity and peak structure among materials and aging conditions reflect variations in crystalline structure.

In addition, the Toyama samples subjected to 50, 70, and 100 aging cycles showed far fewer crystalline peaks and a much more pronounced amorphous halo compared with other compositions, indicating significant structural degradation.

Representative Bayesian peak-deconvolution fits and residuals are shown in Figure 3, confirming the adequacy of the descriptor extraction used for downstream regression.

3.2. Model fit and conservative generalization

Model parameters were determined based on convergence behavior and exchange acceptance ratios. The number of clusters was selected by the estimated free energy at the coldest chain (β = 1), resulting in K = 2 (Fig. S1). Figure 5 shows the actual versus fitted tensile modulus colored by cluster assignment, indicating that the two-cluster model explains the observed dataset (R^2 = 0.81, RMSE = 145 MPa) and supports the existence of cluster-dependent structure-property mappings.

Because tensile tests are destructive and XRD and tensile measurements were not always performed on the same physical specimen, and because multiple XRD profiles share the same tensile label at 0 cycles, we additionally report replicate-holdout group k-fold cross-validation as a conservative generalization estimate (Fig. S2(a), (b)). The tensile-level CV results show reduced performance (R^2 = 0.15, RMSE = 320 MPa, N = 120), which should be interpreted as a lower-bound estimate under specimen-to-specimen variability.

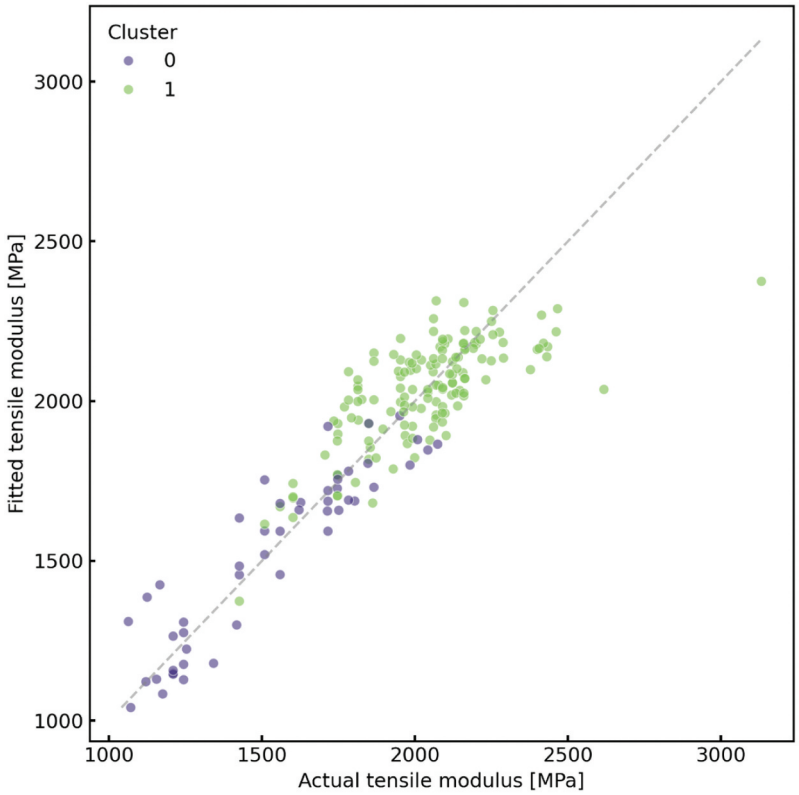


Figure 5. Actual versus fitted (in-sample) tensile modulus colored by cluster assignment. Actual versus fitted tensile modulus for all samples, colored according to cluster assignment inferred by the Bayesian mixture model. The dashed line indicates perfect agreement (identity line ($y = x$)). The model achieves an R^2 value of 0.81 with a root mean squared error of 144.5 MPa.

3.3. Cluster discovery

In this model, two distinct clusters were identified, which can be interpreted as two different structure-property mappings. The distributions of the 2θ position and FWHM of the β(300) peak differed between the two clusters (see Figure 6). In Cluster 0, which corresponds to lower tensile modulus values, the 2θ position tended to be lower and the FWHM also tended to be smaller. In contrast, in Cluster 1, the 2θ position was higher and FWHM distribution exhibited a tail toward larger values. These results indicate that β-phase peak descriptors carry information associated with cluster assignment beyond β fraction alone. However, because Figure 6 summarizes differences in descriptor space, it does not by itself establish that within-cluster regression relationships differ for these descriptors. As shown in Figure 7, polymer composition also influenced cluster assignment. Direct comparison of cluster-specific posterior coefficient distributions for selected β-phase descriptors is presented in Direct comparison of cluster-specific coefficient posteriors.

3.4. Feature contributions

To investigate which XRD-derived features contributed to the prediction of tensile modulus, two heatmaps describing the posterior feature-selection probabilities for each cluster are shown in Figure 8(a) and (b) as visualizations of PIPs. The figures summarize how frequently each descriptor is selected within the regression model of a given cluster and therefore reflect cluster-specific variable-selection tendencies rather than direct between-cluster differences in regression coefficients. The coldest chain ($\beta = 1$) represents the target posterior distribution and is therefore of primary interest. At $\beta = 1$, the heights of the β(300) and α(130) peaks were selected in both clusters. Cluster 0 additionally showed relatively high inclusion probabilities for β (300) peak position, FWHM of α(040), β crystallinity, and overall crystallinity, whereas Cluster 1 preferentially selected α(130) area ratio, halo-region peak height, FWHM of α-related peaks, and overall crystallinity. Notably, β(300) FWHM was selected in both clusters, suggesting shared predictive relevance. We therefore next examine whether the coefficient distributions themselves differ across clusters for selected β-phase descriptors.

3.5. Direct comparison of cluster-specific coefficient posteriors

To directly assess whether β-phase descriptors exhibit cluster-dependent structure–modulus relationships, Figure 9 compares the posterior coefficient distributions conditioned on feature inclusion, $w_{kd} \mid c_{kd} = 1$,

at the coldest chain. Among the β-phase descriptors examined, β(300) peak position showed the clearest between-cluster shift in posterior mass: Cluster 0 favored more positive coefficients, whereas Cluster 1 showed a broader distribution extending toward lower and negative values.

Quantitatively, at the coldest chain ($\beta = 1$), the conditional posterior mean for β(300) peak position was 0.51 in Cluster 0 and −0.02 in Cluster 1, with $P(w_{kd} > 0 | c_{kd} = 1) = 0.79 and 0.36$, respectively, whereas the corresponding values for β(300) FWHM were both 0.34, with $P(w_{kd} > 0 | c_{kd} = 1) = 0.95 and 0.93$, indicating shared positive relevance across clusters. β crystallinity also showed a negative tendency in both clusters, somewhat stronger in Cluster 1, although with broader uncertainty and only moderate inclusion probability. Together, Figures 6, 8, and 9 distinguish three complementary aspects of the model: descriptor-space separation, cluster-specific variable-selection tendencies, and direct differences in regression behavior. Additional conditional coefficient posteriors for other top-ranked descriptors are provided in Figure S3.

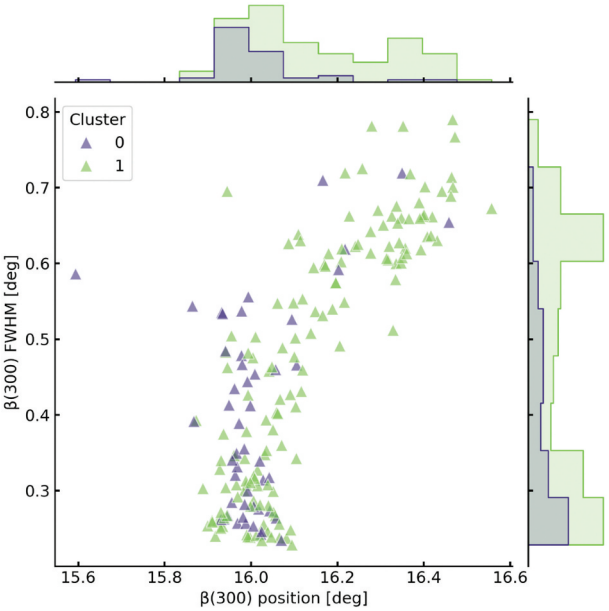


Figure 6. Cluster-wise distributions of β(300) peak position and FWHM. Joint scatter and marginal histograms of β(300) peak position (2θ) and FWHM, colored by inferred cluster assignment. The clusters show a clearer shift in β(300) peak position than in FWHM, indicating cluster-wise differences in descriptor space. This figure summarizes descriptor-space differences only; evidence for cluster-dependent regression behavior is provided separately in Figure 9.

4. Discussion

4.1. Materials science interpretation of the result

4.1.1. Cluster 0

In Cluster 0, not only the amount of β crystal but also β(300)-related peak-shape/position descriptors appeared to influence tensile modulus. These descriptors may reflect variations in the β-phase lattice state (e.g. microstrain, crystallite size, and/or imperfect ordering). Direct comparison of the posterior coefficient distributions indicates that higher β(300) peak position is more consistently associated with higher tensile modulus in this cluster, whereas β(300) FWHM shows a mainly positive tendency in both clusters. Together, these results suggest that β-phase structural descriptors, especially β(300) peak position, can be more informative than β fraction alone for explaining the modulus variability within Cluster 0.

4.1.2. Cluster 1

In Cluster 1, the contribution of the amorphous halo and the relative amount of α crystallites played a dominant role. The frequent selection of the FWHM of α-related peaks indicates that crystalline distortion within the α phase also affected the tensile

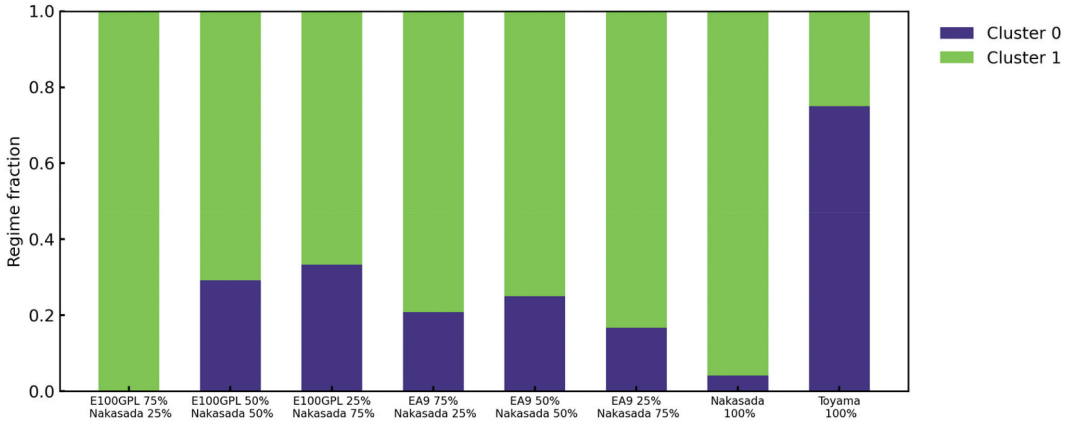


Figure 7. Cluster fractions and polymer composition. Stacked bar plots show the fraction of samples assigned to each cluster (Cluster 0 and Cluster 1) for different virgin/recycled PP compositions. For E100-GPL-based composites, an increasing recycled PP content leads to a higher assignment probability to Cluster 0, indicating a composition-dependent shift in the dominant structure-property cluster.

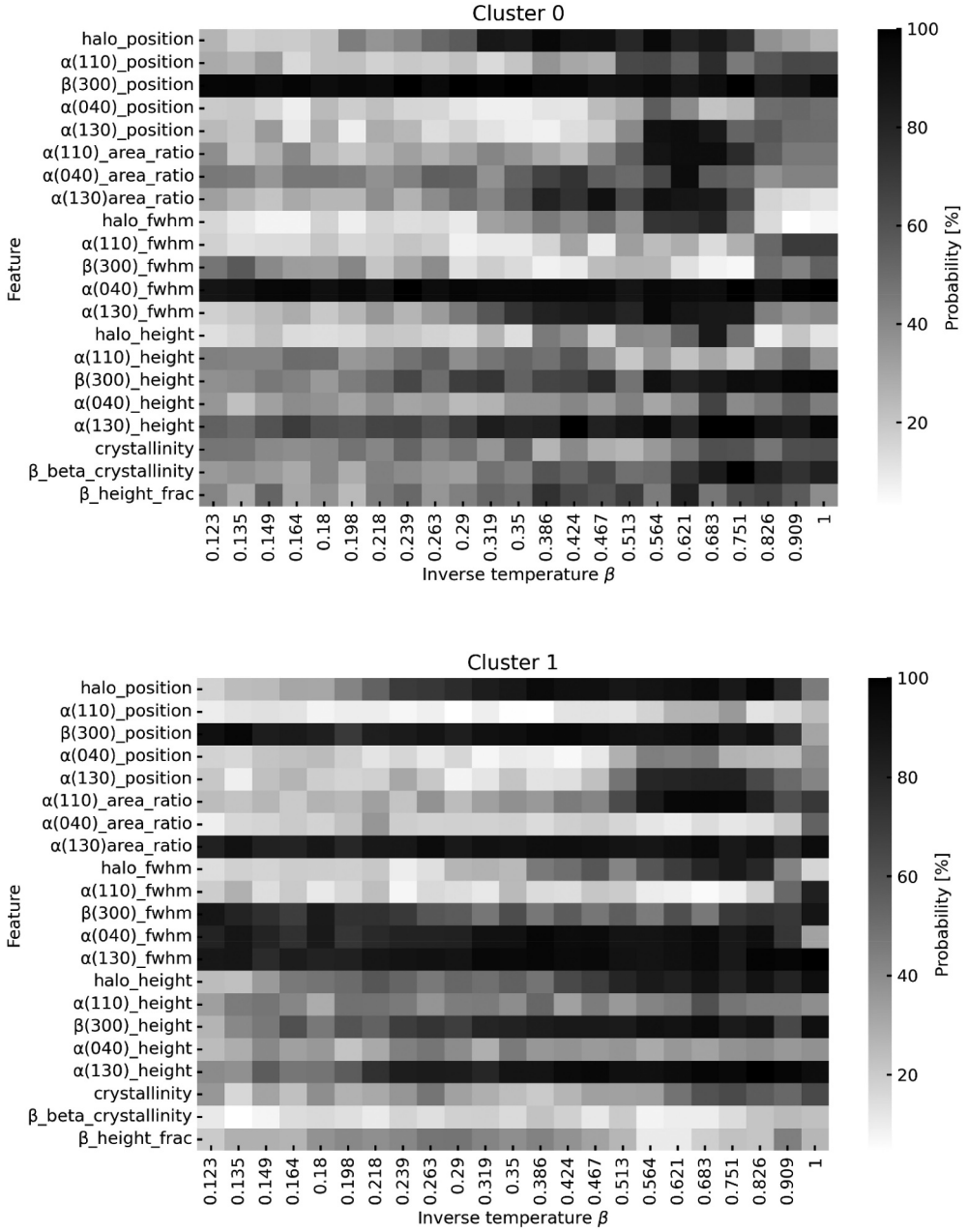


Figure 8. Posterior feature-selection probabilities across inverse temperatures. Heatmaps show the posterior inclusion probabilities of XRD-derived features as a function of inverse temperature β for (a) Cluster 0 and (b) Cluster 1. The probabilities were obtained by averaging binary feature-selection indicators over posterior samples at each temperature, weighted by the occurrence probability of each selection pattern. The coldest chain (β = 1) represents the target posterior distribution and is therefore of primary interest.

modulus in this cluster. Compared with Cluster 0, the posterior for β(300) peak position was broader and less consistently positive, whereas β(300) FWHM retained a mainly positive tendency. Thus, within the β-phase descriptors, the clearest cluster-dependent difference is observed for β(300) peak position rather than for broadening alone.

4.1.3. Cluster-composition relationship

As shown in Figure 7, for composites based on E100-GPL, the probability of assignment to Cluster 0 increased with increasing recycled PP content. In contrast, for EA9-based composites, no clear trend was observed. Figure 6 further shows cluster-wise differences in β(300) peak descriptors, particularly in peak position. Because this figure summarizes differences in descriptor space rather than regression effects, the evidence for a cluster-dependent structure–modulus relationship comes mainly from the direct coefficient comparison in Figure 9. Taken together, these results suggest that composition influences not only cluster prevalence but also the structural regime in which β-phase descriptors contribute to tensile modulus. In particular, β(300) peak position provides the clearest evidence of a cluster-dependent structure–modulus relationship, whereas β(300) FWHM appears relevant in both clusters.

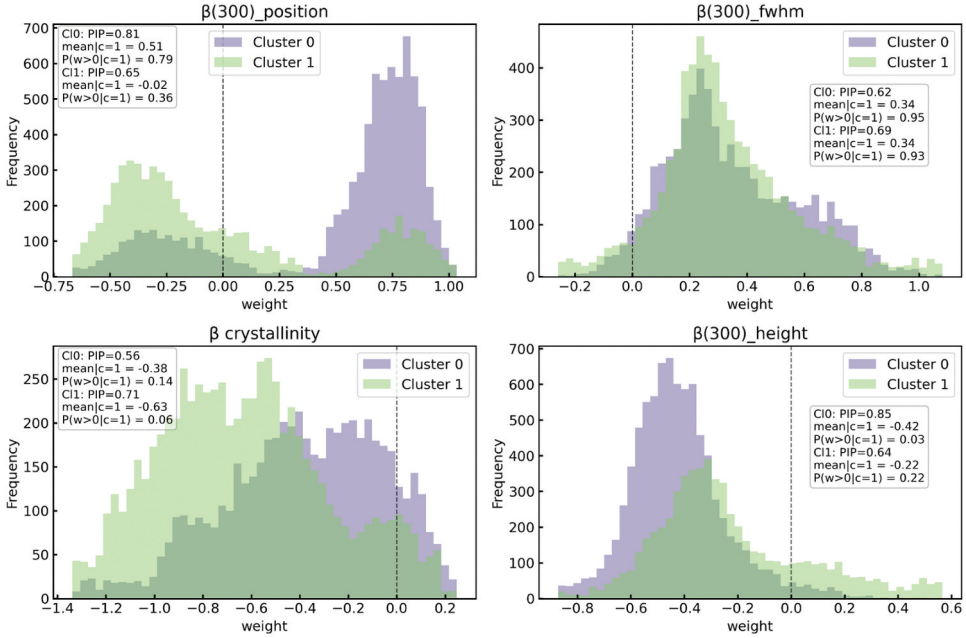


Figure 9. Direct comparison of cluster-specific posterior coefficient distributions for selected β-phase descriptors. Histograms show posterior samples of the standardized regression coefficient w_{kd} conditional on feature inclusion ($c_{kd} = 1$) for Cluster 0 and Cluster 1 at the coldest chain ($\beta = 1$). Panels correspond to β(300) peak position, β(300) FWHM, β crystallinity, and β(300) peak height. Insets report the posterior inclusion probability (PIP), the conditional posterior mean $E[w_{kd} \mid c_{kd} = 1]$, and $P(w_{kd} > 0 \mid c_{kd} = 1)$. Posterior samples were aligned to the optimal cluster ordering before summarization to mitigate label switching. Positive w_{kd} indicates that a larger standardized descriptor value is associated with a higher tensile modulus within the corresponding cluster.

4.1.4. Xenon-arc weathering: dual effects of chemi-crystallization and degradation

Xenon-arc lamp exposure is known to accelerate two competing processes in polypropylene: chain scission followed by molecular reorganization, which can increase crystallinity and stiffen the material, and oxidation or surface damage, which can reduce mechanical strength, particularly at higher exposure cycles. In terms of XRD-derived features, the hardening effect is consistent with increased crystalline order, whereas degradation is directly associated with larger FWHM values and a more pronounced amorphous halo. Importantly, the observed behavior could not be explained solely as a monotonic function of weathering cycle number, indicating that the PP samples followed different structural evolution pathways. This observation motivated the adoption of a cluster-dependent structure-property mapping, rather than a single global relationship, to describe the mechanical behavior of recycled PP.

4.2. Practical implications for grading recycled PP

The primary practical value of the proposed framework is interpretable grading rather than maximizing point prediction accuracy. From an XRD profile, Bayesian peak deconvolution yields a compact set of physically interpretable descriptors, and the mixture model provides (i) cluster-dependent structure–modulus mappings, (ii) cluster-specific important descriptors quantified by PIPs, and (iii) uncertainty-aware modulus estimates. In practice, this enables a grading workflow in which rPP batches are characterized by both the estimated modulus range and the dominant cluster (e.g. a cluster showing stronger positive dependence on β(300) peak position versus a cluster more influenced by α/amorphous descriptors), offering richer diagnostics than a single crystallinity metric. The conservative replicate-holdout CV results further indicate that deployment will benefit from paired same-specimen measurements and/or improved cluster gating when tensile testing is not available.

4.3. Limitations and future work

Several limitations of this study should be acknowledged. First, XRD profiles and tensile modulus were obtained from different specimens, and, for the unaged condition, multiple pre-weathering XRD-derived feature vectors were paired with the same 0-cycle tensile modulus value. This mismatch likely introduces specimen-to-specimen variability and label noise, which can increase residual errors and make the replicate-holdout group k-fold cross-validation a conservative (lower-bound) estimate of generalization performance.

Second, the present study uses a two-stage workflow in which Bayesian peak deconvolution is used for feature extraction and one point-estimate descriptor vector per profile is then passed to the mixture regression model. Although the deconvolution stage provides posterior uncertainty for the peak parameters that full uncertainty

is not propagated into the regression stage in the current implementation. A fully joint hierarchical model that propagates profile-level uncertainty into property prediction would therefore be a valuable extension.

Third, the current model does not explicitly incorporate oxidation chemistry or surface roughness, both of which can significantly influence mechanical properties, particularly under weathering conditions. Fourth, the robustness of the conclusions should be strengthened through validation using external datasets with broader material variability.

Future work will focus on integrating complementary analytical techniques (such as IR or Raman spectroscopy) to capture chemical degradation effects, expanding the dataset to include additional recycled PP sources, and refining the modeling framework, including uncertainty propagation across stages, to further enhance predictive accuracy and industrial applicability.

5. Conclusion

We presented an interpretable Bayesian workflow for nondestructive grading of recycled polypropylene, coupling Bayesian peak deconvolution of XRD profiles with a finite mixture of linear regressions equipped with probabilistic feature selection and REMC. Using 21 physically interpretable descriptors extracted from α/β crystalline peaks and the amorphous halo, the model identified two clusters with distinct structure–modulus mappings and achieved a good in-sample fit ($R^2 = 0.81$). Conservative replicate-holdout validation at the tensile level yielded lower performance ($R^2 = 0.15$), consistent with specimen-to-specimen variability and repeated labels at 0 cycles. Posterior inclusion probabilities indicated that β-phase descriptors carry information beyond β fraction alone. Direct comparison of cluster-specific posterior coefficient distributions further showed that β(300) peak position provides the clearest evidence of cluster-dependent regression behavior, whereas β(300) FWHM remains relevant in both clusters and is therefore better interpreted as a shared predictor than as a uniquely cluster-separating descriptor. These findings refine the interpretation of the β phase in recycled PP and support the use of uncertainty-aware XRD-based analysis for grading heterogeneous rPP streams. Future improvements will benefit from paired same-specimen XRD–tensile measurements and complementary chemical probes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) [Grant Number 23K26724], M. N. acknowledges the support of JST CREST (JPMJCR19J3), MEXT Program: Data Creation and Utilization-Type Material Research and Development Project [Grant Number JPMXP1122714694], and Cross-ministerial Strategic Innovation Promotion Program (SIP) JPJ012290.

ORCID

Kazuki Hammura http://orcid.org/0009-0000-9671-5059
Kenji Nagata http://orcid.org/0000-0001-9894-4461
Masanobu Naito http://orcid.org/0000-0001-7198-819X

References

[1] Geyer R, Jambeck JR, Law KL. Production, use, and fate of all plastics ever made. Sci Adv. 2017;3(7): e1700782. doi: 10.1126/sciadv.1700782
[2] Ragaert K, Delva L, Geem KV. Mechanical and chemical recycling of solid plastic waste. Waste Manag. 2017;69:24–58. doi: 10.1016/j.wasman.2017.07.044
[3] Gall M, Freudenthaler PJ, Fischer J, Lang RW. Characterization of composition and structure-property relationships of commercial post-consumer polyethylene and polypropylene recyclates. Polym. 2021;13(10):1574. doi: 10.3390/polym13101574
[4] Ogawa H, Shibasaki K, Kawai T, et al. Poly(ethylene terephthalate)/aluminum inclusions as critical defects in recycled polypropylene derived from municipal waste: synchrotron-based visualization and fracture analysis. ACS Sustain Resour Manage Forthcom. [cited 2026 Jan 5]. doi: 10.1021/acssusresmgt.5c00526
[5] Plastics the fast facts 2025 [Internet]. Brussels Belgium: Plastics Europe AISBL; [cited 2026 Jan 2]. Available from: https://plasticseurope.org/knowledge-hub/plastics-the-fast-facts-2025/
[6] Varga J. β-modification of isotactic polypropylene: preparation, structure, processing, properties, and application. J Macromol Sci B. 2002;41(4–6):1121–-1171. doi: 10.1081/MB-120013089
[7] Menyhárd A, Suba P, László Z, et al. Direct correlation between modulus and the crystalline structure in isotactic polypropylene. Express Polym Lett. 2015;9(3):308–320. doi: 10.3144/expresspolymlett.2015.28
[8] Shi Y, Rabin T, Mark C, et al. Mechanical properties of recycled plastic fibres for reinforcing concrete. Prague, Czech Republic: Fibre Concrete; 2013 [2013 Sep 12–13].
[9] Yao S, Yominaga Y, Fujikawa Y, et al. Inner structure and mechanical properties of recycled polypropylene. Nihon Reoroji Gakkaishi. 2013;41(3):173–178. doi: 10.1678/rheology.41.173
[10] Rabello MS, White JR. Crystallization and melting behavior of photodegraded polypropylene – i. chemi-crystallization. Polym. 1997;38(26):6379–6387. doi: 10.1016/S0032-3861(97)00213-9
[11] Craig IH, White JR, Kin PC. Crystallization and chemi-crystallization of recycled photo-degraded

polypropylene. Polym. 2005;46(2):505–512. doi: 10.1016/j.polymer.2004.11.019
[12] Shinzawa H, Watanabe R, Yamane S, et al. Aging of polypropylene probed by near infrared spectroscopy. J Near Infrared Spectrosc. 2021;29(5):259–268. doi: 10.1177/0967033521999115
[13] Fernandes C. Scientific machine learning for polymeric materials. Polym. 2025;17(16):2222. doi: 10.3390/polym17162222
[14] Wold S, Ruhe A, Wold H, et al. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J Sci Comput. 1984;5(3):735–743. doi: 10.1137/0905052
[15] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: COLT '92: Proceedings of the fifth annual workshop on Computational learning theory; Pittsburgh (PA) USA. Association for Computing Machinery; 1992 [1992 Jul 27–29].
[16] Thomas AJ, Barocio E, Bilionis I, et al. Bayesian inference of fiber orientation and polymer properties in short fiber-reinforced polymer composites. Compos Sci Technol. 2022;228:109630. doi: 10.1016/j.compscitech.2022.109630
[17] Merrill JH, Han Y, Roth CB. A Bayesian inference approach to accurately fitting the glass transition temperature in thin polymer films. Macromolecules. 2024;57(23):11055–11074. doi: 10.1021/acs.macromol.4c01867
[18] Valente D, Yip HC, Mastrantonio G, et al. Bayesian inference for latent spectral shapes. J R Stat Soc. 2025:1–19. doi: 10.1093/jrsssc/qlaf057
[19] Wolfman M, Di ZW, Ranjan R, et al. Kβ x-ray emission spectra analysis using Bayesian optimization. Photon Sci. 2025. doi: 10.1021/photonsci.5c00007
[20] Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Jpn. 1995;65(6):1604–1608. doi: 10.1143/JPSJ.65.1604
[21] Japanese industrial standards. JIS K 7139 plastics - test specimens. 2009.
[22] Japanese industrial standards. JIS K 7350 - 2 plastics - methods of exposure to laboratory light sources – part 2: xenon-arc lamps. 2008.
[23] Nagata K, Sugita S, Okada M. Bayesian spectral deconvolution with the exchange Monte Carlo method. Neural Netw. 2012;28:82–89. doi: 10.1016/j.neunet.2011.12.001
[24] Tamura R, Nagata K, Sodeyama K, et al. Machine learning prediction of the mechanical properties of injection-molded polypropylene through x-ray diffraction analysis. Sci Technol Adv Mater. 2024;25(1). doi: 10.1080/14686996.2024.2388016
[25] Jones AT, Aizlewood JM, Beckett DR. Crystalline forms of isotactic polypropylene. Macromol Chem Phys. 1963;75(1):134–158. doi: 10.1002/macp.1964.020750113