*Proceedings of the 11th North American International Conference on Industrial Engineering and Operations Management, Milwaukee, Wisconsin, USA, June 9-11, 2026*

*Proceedings of the International Conference on Industrial Engineering and Operations Management*

Publisher: IEOM Society International, USA DOI: 10.46254/NA11.20260406
Published: June 9, 2026

# Dynamic ARIMA Mixture Modeling (DARIMA-MM) for Forecasting Time-Varying Nanoparticle Shape Distributions

**Rifat Hasan and Arda Vanli**
Industrial and Manufacturing Engineering
FAMU-FSU College of Engineering
Florida Agricultural & Mechanical University and Florida State University
Tallahassee, FL, USA
rh20du@fsu.edu, oavanli@eng.famu.fsu.edu

**Chiwoo Park**
Industrial and Systems Engineering
University of Washington
Seattle, WA, USA
chiwpark@uw.edu

## Abstract

Geometrical forms of products are important quality characteristics in advanced manufacturing, e.g., geometries of 3D printed products and morphology of nano-manufactured products. In many material-processing systems, particle populations are heterogeneous, multi-modal and their shape distributions may evolve over time due to uncertain physical and environmental conditions. Hence, tracking the temporal behavior of shape distributions is important for process monitoring, material characterization, and data-driven decision-making. Here, we propose a dynamic mixture-modeling framework for time-varying shape distributions, where particle shapes are represented through clusters and the corresponding mixing proportions are modeled as time-dependent quantities. In this study, we consider that only mixture proportions will change along with time; the other parameters like mean shape and covariance will be constant. Finally, the proposed approach is evaluated using nanoparticle shape-distribution data observed over 20 timepoints.

## Keywords
Dynamic mixture model, ARIMA, shape analysis, nanoparticle characterization, time-series forecasting

## 1. Introduction
In many advanced manufacturing and materials-processing systems, particles are generated and transformed under uncertain physical and environmental conditions. Even when particles are produced under similar process settings, their shapes and sizes may evolve differently over time. As a result, the population of material objects is often better described by a distribution of shapes rather than by a single representative geometry. Tracking how this distribution changes over time is important for understanding material evolution, improving process monitoring, and supporting data-driven decision-making in manufacturing applications (Park and Ding 2021; Ulusoy 2023; Feltner et al. 2025).

Shape distribution tracking focuses on representing probability distributions over object shapes and modeling how those distributions evolve over time. Existing studies have introduced several frameworks for static and dynamic shape modeling, including landmark-based representations (Kendall 1984; Bookstein 1986; Dryden and Mardia 2016), tangent-coordinate models (Fletcher et al. 2004; Klingenberg 2020), offset normal distributions (Dryden and Mardia 1991; Kume and Welling 2010; Fontanella et al. 2019), and curve-based stochastic growth models ( Jonsdottir and Jensen 2005; Park and Ding 2021). These approaches provide useful tools for describing shape variability and its temporal evolution. However, many real material systems contain heterogeneous and multi-modal particle populations, where different shape groups may be present simultaneously and their relative proportions may change across time. Recent studies in nanoparticle image analysis and morphology characterization further highlight the need to identify and quantify multiple particle-shape groups within large material populations (Kozikowski 2023; Glaubitz et al. 2024; Monteiro et al. 2025). These studies signify a need for modeling frameworks that can capture both shape-based heterogeneity, multi-modality and temporal dependence.

Mixture models provide a natural framework for representing heterogeneous populations because they allow observations to arise from multiple latent components ((McLachlan and Peel 2000; Frühwirth-Schnatter and Malsiner-Walli 2019). In the context of particle shape analysis, each component may represent a distinct shape group, while the mixing proportions describe the relative prevalence of those groups in the population. When these proportions vary over time, a static mixture model may not fully capture the underlying dynamics. Time-series models such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models are widely used for temporal modeling (Hamilton 1994), but conventional forms are not designed to directly represent evolving mixture structures or time-varying subpopulation behavior.

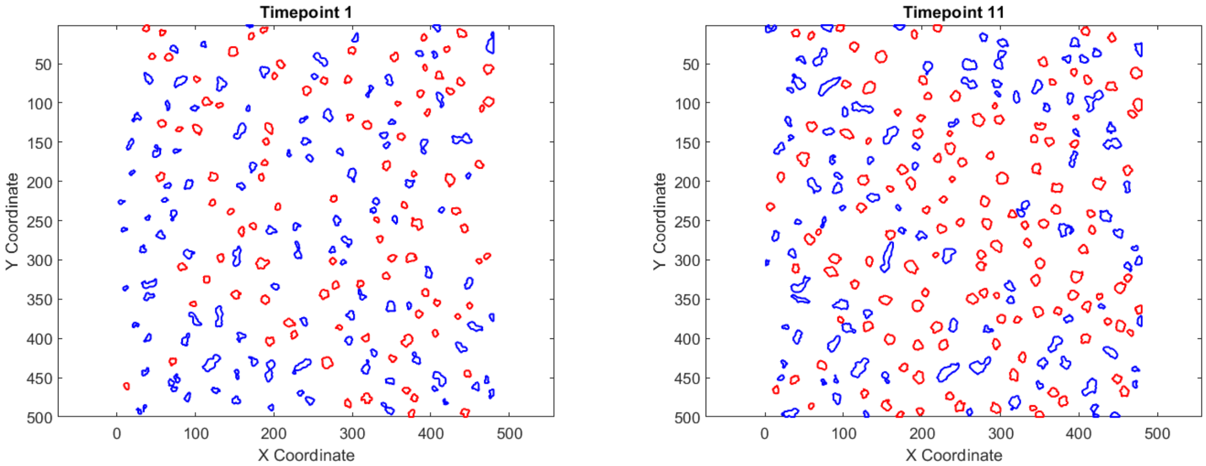
Hidden Markov models (HMMs) provide closely related frameworks for modeling temporal dependence in heterogeneous sequential data. In an HMM, the observed time series is assumed to be generated from an underlying discrete-valued latent state process that evolves according to a Markov transition structure. Conditional on the current hidden state, the observation is generated from a state-dependent emission distribution. Therefore, HMMs can be interpreted as Markov-dependent mixture models in which the mixture component at each time point is governed by an unobserved Markov chain. This structure makes HMMs useful for modeling regime-switching behavior, latent state transitions, and time-dependent subpopulation membership in sequential data (Zucchini et al. 2017; McClintock et al. 2020; Glennie et al. 2023).

Several extensions of HMMs have been developed for heterogeneous and multivariate time-series data. Mixture hidden Markov models allow different groups of sequences to have distinct latent-state dynamics, while coupled and multivariate HMMs can represent interacting latent processes. These models are particularly useful when the main scientific objective is to infer discrete latent regimes or transitions among unobserved states. Recent work has shown that HMM-based mixture structures can improve the modeling of noisy, irregularly sampled, or heterogeneous multivariate time series by combining latent-state dynamics with subgroup-specific behavior (Poyraz and Marttinen 2023). However, HMM-based approaches usually focus on latent state occupancy and transition probabilities, rather than directly forecasting smoothly evolving cluster-specific mixing proportions.

State-space models provide another important class of dynamic latent-variable models. In contrast to standard HMMs, where the latent state is typically discrete, state-space models often use continuous latent variables to describe the evolution of an underlying dynamic process. A general state-space formulation consists of a state equation, which describes how the latent process evolves over time, and an observation equation, which links the latent state to the measured data. Linear Gaussian state-space models are especially attractive because they allow likelihood-based inference and prediction through Kalman filtering and smoothing (Shumway and Stoffer 2017).

These HMM formulations and state-space models are highly relevant to dynamic mixture modeling because they explicitly model temporal dependence through latent dynamic processes. Nevertheless, the proposed Dynamic ARIMA Mixture Model differs in its modeling objective and construction. Rather than assigning each time point to a hidden regime or fitting a state-space model to the full observation sequence, the proposed method first estimates shape-based mixture proportions from particle-shape distributions and then models the temporal evolution of these estimated proportions directly. Thus, the latent shape groups are defined through the shape-mixture model, while the temporal dependence is introduced through ARMA or ARIMA dynamics on the logit-transformed mixing proportions.

To motivate the proposed framework, we consider a nano-particle shape-distribution dataset over 20 timepoints. At each timepoint, particle boundaries were represented as shape observations and initially grouped into two shape-based clusters using $k$-mode clustering (Deng et al. 2022). As an example, Figure 1 illustrates the image data at two representative timepoints used in this study. The particles belong to one of two classes, and the colors of the particle boundaries indicate the known class labels. At Timepoint 1, the mixing proportions are estimated as $[0.46, 0.54]$, while at Timepoint 11 the proportions are estimated to have shifted to $[0.56, 0.44]$. This change suggests that a static mixture model may be insufficient for describing the temporal behavior of the particle population, motivating a dynamic modeling approach for time-varying mixture proportions. In the proposed approach a Gaussian mixture model is fitted through the expectation-maximization (EM) algorithm (Hasan et al. 2026) to the mixture distribution extracted from the image to estimate the relative prevalence of the two shape groups over time (Dempster et al. 1977; McLachlan and Peel 2000). The resulting mixing proportions show that the population structure is not fixed; instead, the proportion of particles assigned to each cluster changes across timepoints.




Mixing Proportion at Timepoint 1 = [0.46, 0.54]

Mixing Proportion at Timepoint 11 = [0.56, 0.44]

Figure 1. Dynamic particle shape distribution (two clusters) through different timepoints. Red and blue boundaries represent particles assigned to two shape-based clusters. The estimated mixing proportions change from $[0.46, 0.54]$ at Timepoint 1 to $[0.56, 0.44]$ at Timepoint 11, indicating temporal variation in the underlying particle population.

To address this limitation, this paper proposes Dynamic ARIMA Mixture Models (DARIMA-MM) for modeling time-varying mixture proportions in shape-distribution data. The proposed framework combines the interpretability of mixture models with the temporal modeling capability of ARMA and ARIMA processes. By modeling the evolution of the mixture proportions over time, the approach can describe how latent shape groups emerge, persist, or decline within a population. This is particularly useful for applications involving nanoparticle characterization, where particle populations may shift dynamically as materials evolve during processing or reaction conditions (Yuan et al., 2021).

The main contribution of this study is the development of a dynamic mixture-modeling framework that links statistical shape analysis with time-series forecasting. The proposed approach provides a flexible way to track time-varying shape distributions, estimate dynamic mixture behavior, and generate forecasts of future population structure. Through real-data applications, the study demonstrates how the proposed models can capture temporal changes in shape distributions more effectively than static mixture models, offering a useful tool for analyzing complex material systems in advanced manufacturing.

## 2. Method

### 2.1 Dynamic Autoregressive Integrated Moving Average Mixture Model (DARIMA-MM)

We consider an image data for a sample of particles that may belong to one of $K$ clusters. Let $\pi_{k,t}$ denote the mixture proportion of the $k$-th cluster $(k = 1,2,\dots,K)$ at time point $t$. This section will derive the ARIMA model for the mixture proportions over multiple points of time in which the means and covariances of the cluster may be non-stationary over time.

First, we derive the expressions for a stationary ARMA model, which consists in autoregressive (AR) and moving average (MA) components. The AR component captures the relationship between an observation and a number of lagged observations, while the MA component represents the relationship between an observation and a residual error from a moving average model applied to lagged observations (Durbin and Koopman 2013; Box et al. 2016). We transform using the logit function the mixture proportions $\pi_{k,t} \in (0,1)$ , to ensure they follow a normal distribution (McCullagh and Nelder 1998). The logit transformation is defined as:

$$\text{logit}(x) = \log\frac{x}{1-x} \tag{1}$$

The logit transformation, which maps the bounded proportion to the real line (Grunwald et al. 1993), that is logit $\pi_{k,t} \in (-\infty, \infty)$. An autoregressive (AR) model of order p on the logit mixture proportion with trend allows to model the dependency of the current mixing proportion at time $t$ on the previous $p$ time-points:

$$\begin{aligned} \mathrm{logit}(\pi_{k,t}) &= \mu_k + \delta_k t + \phi_{k,1}\mathrm{logit}(\pi_{k,t-1}) + \\ &\phi_{k,2}\mathrm{logit}(\pi_{k,t-2}) + \ \dots + \ \phi_{k,p}\mathrm{logit}(\pi_{k,t-p}) + \epsilon_{k,t} \\ &= \mu_k + \delta_k t + \textstyle\sum_{i=1}^{p} \phi_{k,i}\, \mathrm{logit}(\, \pi_{k,t-i}\,) + \epsilon_{k,t} \end{aligned} \tag{2}$$

Here $\phi_{k,\mathrm{i}}$ are the AR coefficients for the $k$-th cluster, and $\epsilon_{k,t}$ is a white noise error term. $\mu_k$ is the intercept term and $\delta_k$ is the linear trend coefficient (McCullagh and Nelder 1998; Box et al. 2016) of the $k$-th cluster that enables modeling linear trends in the logit proportions. The pure AR model can be extended by assuming that the current mixture coefficient depends on the white noise error terms from the previous $q$ time points, which results in an ARMA model of orders p and q:

$$\begin{aligned} logit(\pi_{k,t}) = \ &\mu_k + \delta_k t + \ \phi_{k,1}\ logit(\pi_{k,t-1}) \\ &+ \ \phi_{k,2} logit(\pi_{k,t-2}) + \ \dots \\ &+ \ \phi_{k,p} logit(\pi_{k,t-p\,)} + \epsilon_{k,t} + \ \theta_{k,1}\epsilon_{k,t-1} \\ &+ \ \theta_{k,2}\epsilon_{k,t-2} + \cdots + \ \theta_{k,q}\epsilon_{k,t-q} \\ = \mu_k + \delta_k t + &\sum_{i=1}^{p} \phi_{k,i}\, logit\ (\ \pi_{k,t-i}\ ) + \epsilon_{k,t} \\ &+ \ \sum_{i=1}^{q} \theta_{k,i}\ \epsilon_{k,t-i} \end{aligned} \tag{3}$$

Here $\theta_{k,j}$ are the moving average (MA) parameters for the $k$-th cluster, and $\epsilon_{k,t}$ is a white noise error term assumed to be Normally distributed with 0 mean and variance $\sigma_k^2$. $\epsilon_{k,t}$ can be interpreted as forecast error at time $t$.

Given that mixture proportions are defined for K clusters, we will use a vector ARMA time series in order to include interactions between $K$ clusters (Fletcher et al. 2004). Thus, we modify the equations as follows**.** For each time period $t$, we have a vector of mixture coefficients $\mathrm{logit}\,\uppi_\mathrm{t} = (\mathrm{logit}\,\pi_{1,t}, \mathrm{logit}\,\pi_{2,t}, \dots, \mathrm{logit}\,\pi_{K,t})$. Accordingly, the AR and MA components are represented as matrix equations. Let $\Phi_i$ denote $K \times K$ autoregressive parameter matrices for $i = 1, \dots, p$; let $\Theta_j$ denote $K \times K$ moving average parameter metrices for $j = 1, \dots, q$; let $\epsilon_t$ denote a $K$- dimensional vector of white noise error terms at time $t$. The Vector ARMA Model is given as

$$logit\ (\boldsymbol{\pi_t}) = \boldsymbol{\mu} + \boldsymbol{\delta} t + \sum_{i=1}^{p} \boldsymbol{\Phi_i}\ logit\ (\boldsymbol{\pi_{t-i}}) + \sum_{j=1}^{q} \boldsymbol{\Theta_j}\ \boldsymbol{\epsilon}_{t-j} + \boldsymbol{\epsilon}_t \tag{4}$$

This ARMA model captures both the temporal dependence of the mixture coefficients through the AR component and the residual variation through the MA component. The parameters $\boldsymbol{\Phi_i}$ and $\boldsymbol{\Theta_j}$ can be estimated using maximum likelihood estimation or least squares estimation, as discussed in (Box et al. 2016) and (Hamilton 1994).

In order to capture non-stationary dynamics, the ARMA model will be extended to an Autoregressive-Integrated-Moving Average (ARIMA) model for the logit proportions. Let $y_{k,t} = \ logit\ (\pi_{k,t})$ denote the logit-transformed cluster-specific mixing proportions $\pi_{k,t}$ for cluster $k$. The first-differenced logit series is defined as $\Delta y_{k,t} = y_{k,t} - y_{k,t-1}$. ARIMA (1,1,1) model for the first-differenced logit proportions of the k-th cluster can then be defined as

$$\Delta y_{k,t} = \delta_k + \phi_k \Delta y_{k,t-1} + \theta_k \epsilon_{k,t-1} + \ \epsilon_{k,t} \tag{5}$$

Where $\phi_k$ and $\theta_k$ are the autoregressive and moving-average parameters for cluster $k$, respectively. Note that the intercept term drops when a first difference is taken. To initialize the model at $t = 1$ , we need two prior values so that differencing can be included in the AR structure. We compute the initial innovation as:

$$\epsilon_{k,1} = \ \Delta y_{k,1} - \phi_k\ (\ y_{k,0} - y_{k,-1})$$

Where $y_{k,0} = logit(\pi_{k,0})$ and $y_{k,-1} = logit(\pi_{k,-1})$ and $\pi_{k,0}$ and $\pi_{k,-1}$ are treated as parameters which are also estimated. From the normal density function of the white noise $\epsilon_{k,t} \sim \mathcal{N}(0, \sigma_k^2)$ the likelihood of observing the differenced time series vector $(\Delta y_2, \dots, \Delta y_T)$ is:

$$\begin{aligned} &\mathcal{L}(\phi, \theta, \sigma^2, \delta \mid \pi) \\ &= \prod_{k=1}^{K} \prod_{t=2}^{T} \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}} \exp\left[ -\frac{(\Delta y_{k,t}\ - \delta_k - \ \phi_k \Delta y_{k,t-1} - \ \theta_k \epsilon_{k,t-1})^2}{2\sigma_k^2} \right] \end{aligned} \tag{6}$$

Where $\phi = (\phi_1, \dots \phi_K), \theta = (\theta_1, \dots, \theta_K), \sigma = (\sigma_1, \dots, \sigma_K), \pi = (\pi_1, \dots \pi_K),\ \delta = (\delta_1, \dots, \delta_K)$ are the parameter vectors to be estimated. Maximizing the log likelihood is equivalent to minimizing the negative log likelihood $NLL = -\log \mathcal{L}(\phi, \theta, \sigma^2, \delta \mid \pi)$, which is defined as follows:

$$NLL\ (\phi, \theta, \sigma^2, \delta) = \sum_{k=1}^{K} \sum_{t=2}^{T} \left[ \frac{1}{2} \log(2\pi\sigma_k^2) + \frac{\left(\Delta y_{k,t} - \delta_k - \phi_k \Delta y_{k,t-1} - \theta_k \epsilon_{k,t-1}\right)^2}{2\sigma_k^2} \right] \tag{7}$$

This expression forms the objective function, and we will use it to compute the maximum likelihood estimates (MLE) of the model parameters $\phi, \theta, \delta, \sigma, \pi_0, \pi_{-1}$.

## 2.2 Multi-step Forecasting

After parameter estimation, we apply the ARIMA (1,1,1) model to the logit-transformed cluster-specific mixing proportions in order to obtain forecasted values of mixing proportions $\pi_{k,t}$ for each cluster $k$. To initialize the forecasting procedure, we use the estimated prior mixing proportions $\pi_{k,0}$ and $\pi_{k,-1}$. The initial differenced logit values are $\Delta y_{k,1} = logit\ (\pi_{k,1}) - logit\ (\pi_{k,0})$ and $\Delta y_{k,0} = logit\ (\pi_{k,0}) - logit\ (\pi_{k,-1})$.

For $t = 2, \dots, T$, the one-step ahead forecast of the first-differenced logit proportions will follow ARIMA (1,1,1) forecasting equation:

$$\widehat{\Delta y}_{k,t+1|t} = \phi_k \Delta y_{k,t} + \theta_k \hat{\epsilon}_{k,t} \tag{8}$$

In which an updating equation for the estimated forecast error is defined as

$$\hat{\epsilon}_{k,t} = logit\ (\pi_{k,t}) - logit\ (\pi_{k,t-1}) - \hat{\phi}_k \left[ logit\ (\pi_{k,t}) - logit\ (\pi_{k,t-1}) \right] \tag{9}$$

The corresponding forecasted logit value is then obtained by integrating the forecasted difference:

$$\hat{y}_{k,t+1|t} = y_{k,t} + \widehat{\Delta y}_{k,t|t} \tag{10}$$

Finally, the predicted mixing proportion is obtained by applying the inverse-logit transformation:

$$\hat{\pi}_{k,t+1|t} = \frac{1}{1 + \exp(-\hat{y}_{k,t+1|t})}$$

One-step forecast equation can be extended to determine multi-step forecasts. At forecast origin $t$, the current difference $\Delta y_{k,t}$ and the estimated innovation $\hat{\epsilon}_{k,t}$ are available. To derive a two-step ahead forecast for time period $t + 2$, we write the ARIMA (1,1,1) model for the first-differenced logit proportions of the $k$-th cluster:

$$\Delta y_{k,t+2} = \phi_k \Delta y_{k,t+1} + \theta_k \epsilon_{k,t+1} + \epsilon_{k,t+2} \tag{11}$$

At time $t$, the future innovations are unknown, thus their conditional expectations are zero: $E\left[\epsilon_{k,t+1} \middle| \boldsymbol{\mathcal{F}_t}\right) = 0, E\left[\epsilon_{k,t+2} \middle| \boldsymbol{\mathcal{F}_t}\right) = 0$ , where $\boldsymbol{\mathcal{F}_t}$ denotes the information available up to time $t$. Therefore, a recursive expression to update the one-step ahead forecast Equation (8) to obtain a two-step ahead forecast of the first-differenced logit proportions

$$\widehat{\Delta y}_{k,t+2|t} = \phi_k \widehat{\Delta y}_{k,t+1|t}$$

Similarly, the three step ahead forecasted difference becomes

$$\widehat{\Delta y}_{k,t+3|t} = \phi_k \widehat{\Delta y}_{k,t+2|t} = \phi_k^2 \widehat{\Delta y}_{k,t+1|t}$$

The general expression for $l$- step ahead forecast of the first-differenced logit proportions (for $l = 1,2, \dots, L$) is:

$$\begin{aligned} \widehat{\Delta y}_{k,t+l|t} &= \phi_k^{l-1}\ \widehat{\Delta y}_{k,t+1|t} \\ &= \phi_k^{l-1} \left( \phi_k \Delta y_{k,t} + \theta_k \hat{\epsilon}_{k,t} \right) \end{aligned} \tag{12}$$

Finally, the $l$-step-ahead forecast ($l = 1,2, \dots$,) of logit proportions are:

$$\hat{y}_{k,t+l|t} = y_{k,t} + \sum_{j=1}^{l} \widehat{\Delta y}_{k,t+j|t}$$

$$= y_{k,t} + \left( \phi_k \Delta y_{k,t} + \theta_k \hat{\epsilon}_{k,t} \right) \sum_{j=1}^{l} \phi_k^{j-1}$$

$$= y_{k,t} + \left( \phi_k \Delta y_{k,t} + \theta_k \hat{\epsilon}_{k,t} \right) \frac{1 - \phi_k^l}{1 - \phi_k} \tag{13}$$

In which the last equation is obtained by replacing the summation term with a finite geometric series (assuming $\phi_k \neq 1$) as:

$$\sum_{j=1}^{l} \phi_k^{j-1} = \frac{1 - \phi_k^l}{1 - \phi_k}$$

### 2.3 Forecast Uncertainty

To visualize the uncertainty associated with the ARIMA (1,1,1) forecasts, we construct a fan chart for the predicted cluster-specific mixing proportions. The fan chart displays the future point forecasts together with their corresponding prediction intervals (Wallis 1999; Cogley et al. 2005). ARIMA-based prediction intervals are commonly derived under the assumption that the residual innovations are approximately uncorrelated and normally distributed (Box et al. 2016).
Let $\hat{y}_{k,T+l|T}$ denote the $l$-step-ahead forecast from the final observed time point $T$. The forecast uncertainty in the logit scale can be expressed using the moving-average representation of the ARIMA model. Specifically, the forecast error variance for the $l$-step-ahead forecast is computed as:

$$Var\left( y_{k,T+l} - \hat{y}_{k,T+l|T} \right) = \hat{\sigma}_k^2 \sum_{j=0}^{l-1} \psi_{k,j}^2 \tag{14}$$

Where $\hat{\sigma}_k^2$ is the estimated innovation variance and $\psi_{k,j}$ denotes the $j$-th impulse response coefficient of the integrated ARIMA representation (Box *et al.* 2016; Shumway and Stoffer 2017). For the ARIMA (1,1,1) model, these weights are given by

$$\psi_{k,j} = 1 + (\phi_k + \theta_k) \frac{1 - \phi_k^j}{1 - \phi_k}, \quad j = 0,1, \ldots, l-1$$

Thus, the standard error of the $l$-step ahead forecast in the logit scale is $SE_{k,l} = \sqrt{\hat{\sigma}_k^2 \sum_{j=0}^{l-1} \psi_{k,j}^2}$ . A 95% prediction interval in the logit scale can be computed as $\hat{y}_{k,T+l|T} \pm 1.96\, SE_{k,l}$ .

## 3. Results

In this section, we evaluate the forecasting performance of the proposed Dynamic ARIMA Mixture Model using the estimated time-varying mixing proportions from the nanoparticle shape-distribution data we discussed in Introduction. The dataset consists of 20 points equally spaced in time, where each timepoint contains particle shape observations represented by two shape clusters. We first estimate the cluster-specific mixing proportions using an EM algorithm (Hasan et al. 2026). In this section how to apply the proposed vector ARIMA (1,1,1) model to the cluster-specific logit-transformed mixing proportion series. The results are presented in terms of one-step-ahead forecasting, forecasting error, staggered multi-step forecasting, and prediction uncertainty using a fan chart.

### 3.1 One-Step-Ahead Forecasting and Forecasting Error

Figure 2 compares the observed and one-step-ahead forecasted mixing proportions for the two nanoparticle shape clusters. For Cluster 1, the observed mixing proportion shows an overall increasing pattern across the 20 timepoints, although local fluctuations are present. The ARIMA (1,1,1) forecast follows the general upward trend and captures the overall temporal movement of the cluster proportion. For Cluster 2, the observed mixing proportion shows a decreasing pattern, which is expected because the two cluster proportions are complementary. The corresponding one-step-ahead forecasts also follow this declining behavior.

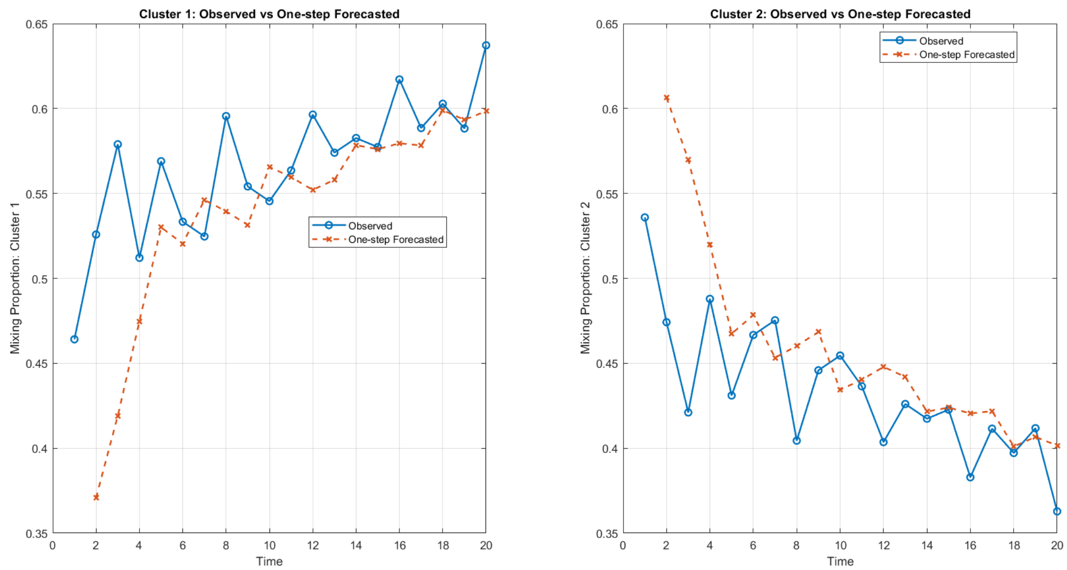



Figure 2. Observed and one-step-ahead forecasted mixing proportions for the two nanoparticle shape clusters using the ARIMA (1,1,1) model.

The forecasts do not exactly match every short-term fluctuation, especially at earlier timepoints where the series changes more sharply. However, the predicted values remain close to the observed proportions for most later timepoints. This indicates that the ARIMA (1,1,1) model is able to capture the dominant temporal structure in the dynamic mixture proportions.

To quantitatively evaluate the one-step-ahead forecasting performance, the root mean squared error (RMSE) was computed for both clusters. The RMSE values are reported in Table 1.

Table 1. One-step-ahead forecasting RMSE for ARIMA (1,1,1).

| Cluster | RMSE |
|---|---|
| Cluster 1 | 0.057377 |
| Cluster 2 | 0.052313 |

The RMSE values indicate that the one-step-ahead forecasts are reasonably close to the observed mixing proportions. Cluster 2 has a slightly smaller RMSE than Cluster 1, suggesting marginally better forecasting accuracy for the second cluster.

### 3.2 Staggered Multi-Step Forecasting and Forecasting Error

Figure 3 shows the staggered multi-step forecasts obtained from the ARIMA (1,1,1) model for Cluster 1. In this analysis, forecasts from one-step to ten-step ahead were generated from different forecast origins. This allows the model's performance to be examined across multiple forecasting horizons.

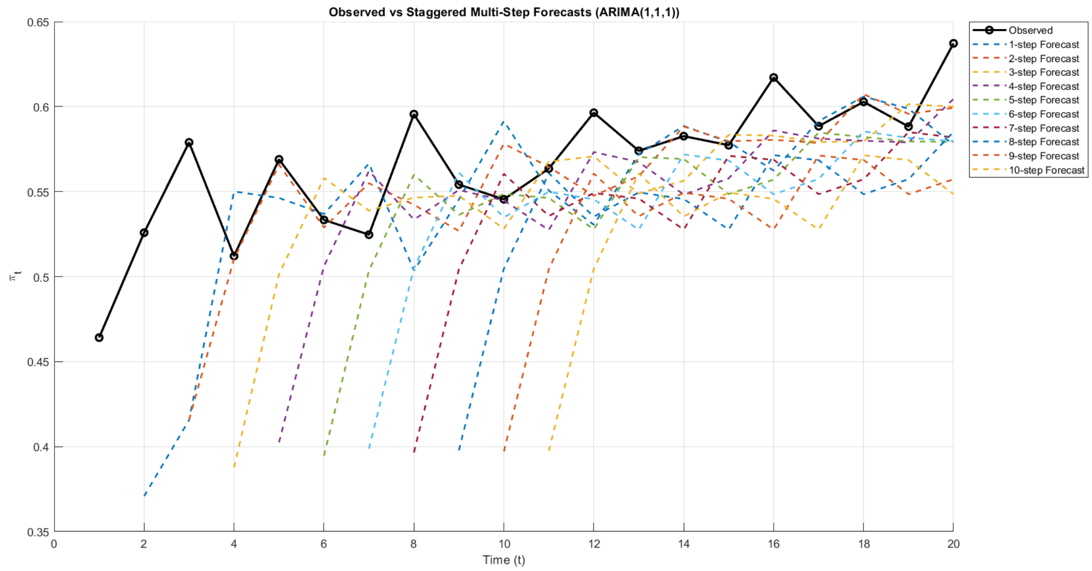



Figure 3. Observed mixing proportions and staggered multi-step forecasts for Cluster 1 using the vector ARIMA (1,1,1) model.

The multi-step forecasts generally follow the central movement of the observed mixing proportion series. As the forecasting horizon increases, the forecast paths become smoother and less responsive to short-term fluctuations. This behavior is expected because future innovation terms are unknown and are replaced by their conditional expectation of zero in multi-step forecasting. Therefore, longer-horizon forecasts depend more heavily on the recursive structure of the ARIMA model than on newly observed shocks.

To further evaluate the effect of forecasting horizon, Figure 4 reports the RMSE values for forecast steps 1 through 10. The RMSE remains relatively low for the early forecast horizons, particularly between steps 2 and 4, indicating that the model provides stable short-term forecasts. After step 5, the RMSE gradually increases, and the largest error is observed at step 10. This pattern suggests that prediction accuracy decreases as the forecasting horizon becomes longer, which is expected in recursive multi-step forecasting because uncertainty accumulates over time.

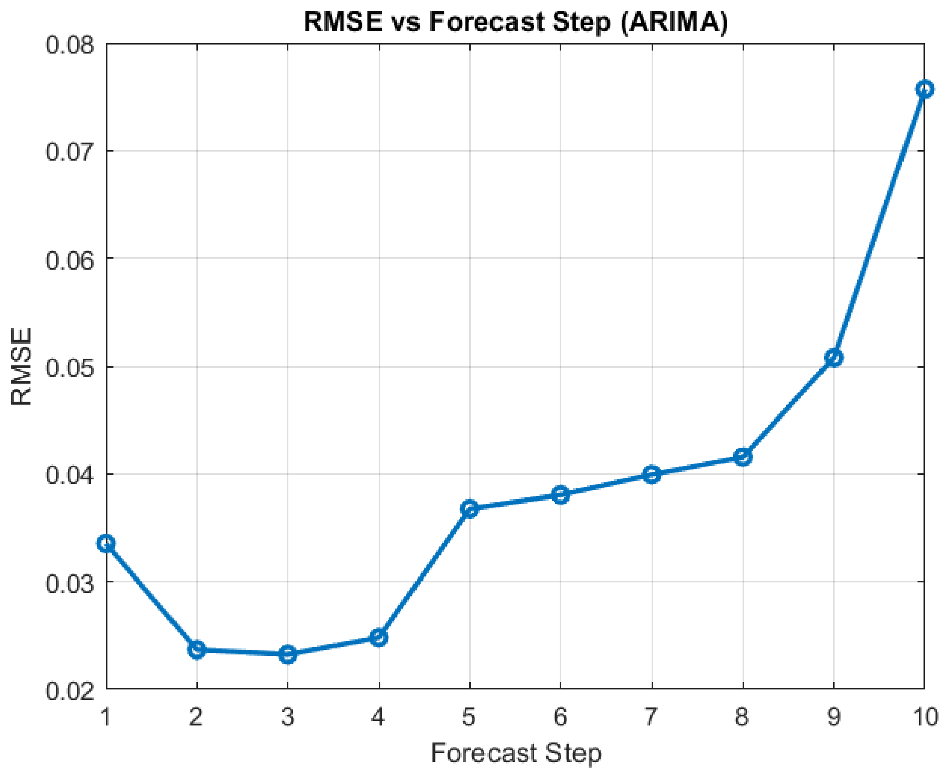



Figure 4. RMSE values across forecast steps for the ARIMA (1,1,1) multi-step forecasting model.

Overall, the staggered forecast paths and horizon-specific RMSE values indicate that the ARIMA (1,1,1) model is more reliable for short- to medium-term forecasting of the nanoparticle shape-cluster proportions. The increasing RMSE at longer horizons highlights the importance of incorporating forecast uncertainty when interpreting future predictions.

### 3.3 Forecast Uncertainty Using Fan Chart

Figure 5 presents the ARIMA (1,1,1) fan chart for Cluster 1. The black line represents the observed mixing proportions up to the final observed timepoint, while the blue line represents the future forecasted mixing proportions. The shaded region represents the 95% prediction interval.

The fan chart shows that the forecasted mixing proportion for Cluster 1 remains relatively stable after the final observed timepoint, with values centered around the recent observed level. At the same time, the prediction interval becomes wider as the forecast horizon increases. This widening interval reflects the accumulation of forecast uncertainty over time. Such behavior is consistent with multi-step ARIMA forecasting, where future predictions depend recursively on previously forecasted values and unobserved future innovations.

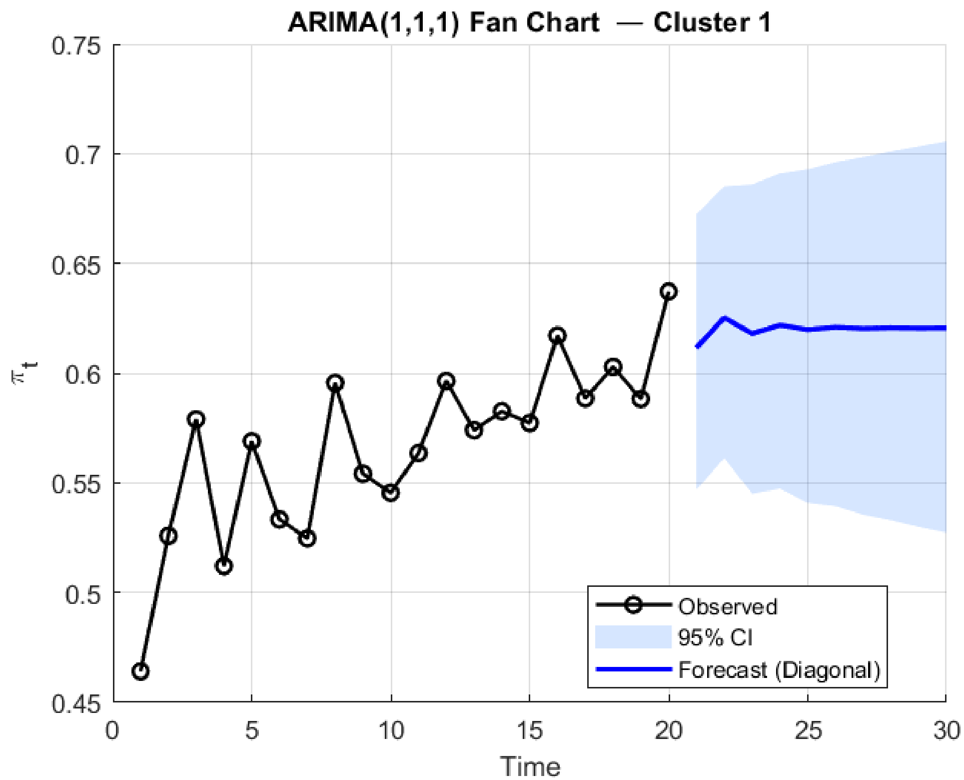



Figure 5. ARIMA (1,1,1) fan chart for Cluster 1 showing the forecasted mixing proportion and 95% prediction interval.

The fan chart is useful because it does not only provide a single future trajectory, but also summarizes the uncertainty associated with the predicted population structure. In the context of nanoparticle characterization, this allows us to describe both the expected future behavior of the shape clusters and the range of plausible variation in their future mixing proportions.

## 4. Discussion

The results show that the nanoparticle shape population exhibits clear time-varying mixture behavior. The proportion of particles associated with each shape cluster changes over time, indicating that a static mixture model is not sufficient to describe the evolving particle population. The ARIMA (1,1,1)-based dynamic mixture model captures the main temporal pattern of the mixing proportions and provides reasonable one-step-ahead forecasting accuracy. The multi-step forecasting and fan chart results further demonstrate that the proposed framework can be used not only to estimate past dynamic behavior, but also to forecast future changes in the particle shape distribution with uncertainty quantification.

These findings support the use of the proposed DARIMA-MM framework as a useful tool for tracking and forecasting time-varying shape distributions in nanoparticle and advanced manufacturing applications.

## 5. Conclusion and Future Work

This study presented a dynamic mixture-modeling framework for tracking and forecasting time-varying nanoparticle shape distributions. The proposed approach combines shape-based mixture modeling with time-series forecasting by first estimating the cluster-specific mixing proportions of nanoparticle shapes and then modeling their temporal evolution using ARMA and ARIMA structures. In this way, the framework provides an interpretable approach for describing how the relative prevalence of different nanoparticle shape groups changes over time.

The real-data study showed that the nanoparticle population exhibits clear time-varying mixture behavior. The estimated proportions of the two shape clusters changed across the 20 observed timepoints, indicating that a static mixture model is not sufficient to describe the evolving particle population. The ARIMA (1,1,1)-based dynamic mixture model captured the main temporal pattern of the mixing proportions and provided reasonable one-step-ahead forecasting performance. The staggered multi-step forecasting results further showed that the model performs more reliably for short- to medium-term forecasting, while forecast error increases at longer horizons. The fan chart analysis provided an additional uncertainty quantification tool by showing both the expected future trajectory and the widening prediction interval over time.

Overall, the results demonstrate that the proposed DARIMA-MM framework can be used to model, forecast, and interpret dynamic changes in nanoparticle shape populations. This framework is particularly useful for advanced manufacturing and materials-processing applications, where the relative prevalence of different particle shape groups may shift over time as the material system evolves.

The current study has several limitations. First, although the proposed approach models the temporal evolution of the mixing proportions, it does not fully represent the entire mixture structure as a dynamic latent process. Conventional ARMA and ARIMA models are effective for modeling temporal dependence in an observed time series, but they are not designed to directly model evolving mixture membership, latent regime switching, or time-varying subpopulation structures. Therefore, related dynamic latent-variable models, such as hidden Markov models and state-space models, represent important methodological competitors and extensions. These models may provide additional flexibility when the underlying particle population evolves through discrete latent regimes or continuous latent state processes. In the present study, the focus was on directly forecasting the cluster-specific mixing proportions estimated from shape-distribution data, rather than modeling latent state transitions.

Second, the current framework assumes that the cluster-specific mean shapes and covariance structures remain fixed over time, while only the mixing proportions evolve. This assumption improves interpretability and allows the temporal behavior of the relative cluster prevalence to be studied directly. However, in some material systems, the shape groups themselves may also evolve over time. Future research will therefore consider models in which the mean shapes, covariance structures, and mixing proportions are all allowed to vary dynamically.

Third, the current real-data application focuses on a two-cluster nanoparticle dataset observed over 20 timepoints. Future studies will evaluate the proposed framework using larger numbers of clusters, longer temporal sequences, and more complex shape populations. This will allow further assessment of the scalability and robustness of the method under more challenging material-characterization settings.

Future work will extend the proposed framework in several directions. In particular, hidden Markov mixture models and state-space mixture models will be investigated as alternative or complementary approaches for capturing more complex temporal behavior in shape-distribution data. Nonlinear time-series models may also be incorporated to improve long-horizon forecasting performance. Finally, the proposed framework will be applied to additional real-time material characterization datasets to further evaluate its usefulness for process monitoring, quality control, and predictive analysis in advanced manufacturing.

## References

Bookstein, F. L., Size and shape spaces for landmark data in two dimensions, *Statistical Science*, vol. 1, no. 2, 1986. doi: 10.1214/ss/1177013696.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M., *Time Series Analysis: Forecasting and Control*, 5th ed., John Wiley & Sons, Hoboken, New Jersey, 2016.

Cogley, T., Morozov, S. and Sargent, T. J., Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system, *Journal of Economic Dynamics and Control*, vol. 29, no. 11, pp. 1893–1925, 2005. doi: 10.1016/j.jedc.2005.06.005.

Dempster, A. P., Laird, N. M. and Rubin, D. B., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 39, no. 1, pp. 1–22, 1977. doi: 10.1111/j.2517-6161.1977.tb01600.x.

Deng, X., et al., Characterizing cell shape distributions using k-mode kernel mixtures, *Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR)*, Montreal, QC, Canada, pp. 2517–2523, 2022. doi: 10.1109/ICPR56361.2022.9956604.

Dryden, I. L. and Mardia, K. V., General shape distributions in a plane, *Advances in Applied Probability*, vol. 23, no. 2, pp. 259–276, 1991. doi: 10.2307/1427747.

Dryden, I. L. and Mardia, K. V., *Statistical Shape Analysis with Applications in R*, 2nd ed., John Wiley & Sons, Chichester, UK, 2016.

Durbin, J. and Koopman, S. J., *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, 2013.

Feltner, L., et al., Particle size and shape analyses for powder bed additive manufacturing, *Particuology*, vol. 101, pp. 33–42, 2025. doi: 10.1016/j.partic.2023.09.001.

Fletcher, P. T., et al., Principal geodesic analysis for the study of nonlinear statistics of shape, *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 995–1005, 2004. doi: 10.1109/TMI.2004.831793.

Fontanella, L., Ippoliti, L. and Kume, A., The offset normal shape distribution for dynamic shape analysis, *Journal of Computational and Graphical Statistics*, vol. 28, no. 2, pp. 374–385, 2019. doi: 10.1080/10618600.2018.1530118.

Frühwirth-Schnatter, S. and Malsiner-Walli, G., From here to infinity: sparse finite versus Dirichlet process mixtures in

model-based clustering, *Advances in Data Analysis and Classification*, vol. 13, no. 1, pp. 33–64, 2019. doi: 10.1007/s11634-018-0329-y.
Glaubitz, C., et al., Leveraging machine learning for size and shape analysis of nanoparticles: A shortcut to electron microscopy, *The Journal of Physical Chemistry C*, vol. 128, no. 1, pp. 421–427, 2024. doi: 10.1021/acs.jpcc.3c05938.
Glennie, R., et al., Hidden Markov models: Pitfalls and opportunities in ecology, *Methods in Ecology and Evolution*, vol. 14, no. 1, pp. 43–56, 2023. doi: 10.1111/2041-210X.13801.
Grunwald, G. K., Raftery, A. E. and Guttorp, P., Time series of continuous proportions, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 55, no. 1, pp. 103–116, 1993. doi: 10.1111/j.2517-6161.1993.tb01470.x.
Hamilton, J. D., *Time Series Analysis*, Princeton University Press, Princeton, NJ, 1994.
Hasan, R., Park, C. and Vanli, A., Statistical modeling of nano-particle shape mixtures, unpublished manuscript, to be submitted, 2026.
Jónsdóttir, K. Ý. and Jensen, E. B. V., Gaussian radial growth, *Image Analysis & Stereology*, vol. 24, no. 2, pp. 117–126, 2011. doi: 10.5566/ias.v24.p117-126.
Kendall, D. G., Shape manifolds, Procrustean metrics, and complex projective spaces, *Bulletin of the London Mathematical Society*, vol. 16, no. 2, pp. 81–121, 1984. doi: 10.1112/blms/16.2.81.
Klingenberg, C. P., Walking on Kendall's shape space: Understanding shape spaces and their coordinate systems, *Evolutionary Biology*, vol. 47, no. 4, pp. 334–352, 2020. doi: 10.1007/s11692-020-09513-x.
Kozikowski, P., Machine learning for grouping nano-objects based on their morphological parameters obtained from SEM analysis, *Micron*, vol. 171, article 103473, 2023. doi: 10.1016/j.micron.2023.103473.
Kume, A. and Welling, M., Maximum likelihood estimation for the offset-normal shape distributions using EM, *Journal of Computational and Graphical Statistics*, vol. 19, no. 3, pp. 702–723, 2010. doi: 10.1198/jcgs.2010.09190.
McClintock, B. T., et al., Uncovering ecological state dynamics with hidden Markov models, *Ecology Letters*, vol. 23, no. 12, pp. 1878–1903, 2020. doi: 10.1111/ele.13610.
McCullagh, P. and Nelder, J. A., *Generalized Linear Models*, 2nd ed., Chapman & Hall/CRC, Boca Raton, 1998.
McLachlan, G. J. and Peel, D., *Finite Mixture Models*, Wiley, New York, 2000.
Monteiro, G. A. A., et al., Pre-trained artificial intelligence-aided analysis of nanoparticles using the segment anything model, *Scientific Reports*, vol. 15, no. 1, article 2341, 2025. doi: 10.1038/s41598-025-86327-x.
Park, C. and Ding, Y., *Data Science for Nano Image Analysis*, Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-030-72822-9.
Poyraz, O. and Marttinen, P., Mixture of coupled HMMs for robust modeling of multivariate healthcare time series, *Proceedings of the 3rd Machine Learning for Health Symposium*, PMLR, vol. 225, pp. 461–479, 2023. Available: https://proceedings.mlr.press/v225/poyraz23a.html, Accessed on June 2, 2026.
Shumway, R. H. and Stoffer, D. S., *Time Series Analysis and Its Applications: With R Examples*, Springer International Publishing, Cham, 2017. doi: 10.1007/978-3-319-52452-8.
Ulusoy, U., A review of particle shape effects on material properties for various engineering applications: From macro to nanoscale, *Minerals*, vol. 13, no. 1, article 91, 2023. doi: 10.3390/min13010091.
Wallis, K. F., Asymmetric density forecasts of inflation and the Bank of England's fan chart, *National Institute Economic Review*, vol. 167, pp. 106–112, 1999. doi: 10.1177/002795019916700111.
Yuan, W., et al., In situ manipulation of the active Au-$TiO_2$ interface with atomic precision during CO oxidation, *Science*, vol. 371, no. 6528, pp. 517–521, 2021. doi: 10.1126/science.abe3558.
Zucchini, W., MacDonald, I. L. and Langrock, R., *Hidden Markov Models for Time Series: An Introduction Using R*, 2nd ed., Chapman and Hall/CRC, Boca Raton, 2017. doi: 10.1201/b20790.

## Biographies

**Rifat Hasan** is a Ph.D. candidate in the Department of Industrial and Manufacturing Engineering at the FAMU- FSU College of Engineering. His research interest is in Machine Learning and Data science with applications to nanoparticle characterization in advanced manufacturing. His research focuses on developing computational and statistical methods for image and shape analysis.

**Dr. Arda Vanli** is a Professor in the Industrial and Manufacturing Engineering (IME) Department at Florida A&M University and Florida State University, College of Engineering. Dr. Vanli's research interests lie in the general area of applied industrial statistics and data analytics with applications in quality and reliability improvement in modern manufacturing processes, risk and vulnerability analysis for natural disasters and infectious disease data analysis. His research is published in journals including Quality Engineering, Quality and Reliability Engineering International, Technometrics, IIE Transactions, IEEE Transactions on Semiconductor Manufacturing and Mechanical Systems and Signal Processing. He completed his Ph.D. in Industrial Engineering and Operations Research at the Pennsylvania State

University, University Park, in August 2007. He received his M.S. degree in Mechanical Engineering from the Pennsylvania State University, University Park, PA in 2000 and his B.S. degree in Mechanical Engineering from the Middle East Technical University, Ankara, Turkey in 1998.

**Dr. Chiwoo Park** is a Professor in the Department of Industrial and Systems Engineering at University of Washington. His main research interest lies in machine learning and data science with applications to advanced manufacturing and physical science. He is particularly interested in modeling and analysis of object data. It concerns statistical analysis of complex objects and their visual features (such as image, shape, motion, function and directions). Many of his methodological studies are motivated by the problem of understanding processing-structure-property relations in manufacturing and physical sciences. His research has been published in journals including IISE Transactions, Operations Research, Journal of Machine Learning Research, Pattern Recognition, IEEE Transactions on Medical Imaging, and Journal of Computational and Graphical Statistics. In 2021, he authored the book Data Science for Nano Image Analysis, which summarizes many of his works in these areas. Another application is Data Science for Motion and Time Analysis in Operations Research. He obtained B.S. degree in Industrial Engineering from Seoul National University in 2001. He received Ph.D. degree in Industrial Engineering from Texas A&M University, College Station, TX in 2011 under the guidance of Yu Ding in Industrial and Systems Engineering and Jianhua Huang in Statistics at Texas A&M University. Before joining the University of Washington in 2024, he was a Professor in the Department of Industrial and Manufacturing Engineering at Florida State University.