Automatic Gaussian Mixture Sampling for Uncertainty Management in Circular Factories

Zahra Sadat Hajseyed Nasrollah[1,2], Ali Darijani[1,2], Mehdi Khabou[3], Stefan Dietrich[3], Volker Schulze[3], and Jürgen Beyerer[1,2]

[1] Fraunhofer IOSB, Karlsruhe BW 76131, Germany
{zahra.sadat.hajseyed.nasrollah,ali.darijani,juergen.beyerer}@iosb.fraunhofer.de
[2] KIT-IES, Karlsruhe BW 76131, Germany
{zahra.nasrollah,ali.darijani,juergen.beyerer}@kit.edu
[3] KIT-IAM, Karlsruhe BW 76131, Germany
{mehdi.khabou,stefan.dietrich,Volker.Schulze}@kit.edu

Abstract. Returned components in circular manufacturing often form heterogeneous condition populations that are poorly represented by a single Gaussian distribution. Such parts arrive with different usage, inspection, repair, and degradation histories. As a result, product condition, process duration, and quality risk are better viewed as random variables with several condition regimes rather than as one homogeneous population. In this work, we represent this structure using Gaussian mixture models (GMMs), which combine discrete regime information and continuous variability in one probabilistic description. Many quantities needed for planning and risk assessment do not admit tractable closed-form expressions once GMM uncertainties are propagated through nonlinear engineering models or decision rules. We therefore present an automatic Monte Carlo sampling layer for GMM-based uncertainty propagation. The user specifies a tolerance and confidence level, and sampling continues until the estimated confidence interval for the selected output quantity satisfies the requested precision or the maximum sampling budget is reached. The approach enables GMM-based uncertainty models to be used in simulation-based planning, routing, inspection, and quality assessment in circular factories without manually fixing the Monte Carlo sample size in advance. We demonstrate the workflow with a proof of concept fatigue strength example for returned or treated metallic components. Multimodal surface state uncertainty is propagated through a nonlinear fatigue strength relation and the output is assessed by its moments and by a low strength probability.

Keywords: Gaussian mixture models · Monte Carlo sampling · uncertainty propagation · circular factory · fatigue strength

1 Motivation and Scope

Circular production aims to keep product value in the system by returning, inspecting, updating, repairing, and reintegrating product instances into new

production cycles [7,8,11,13]. In contrast to purely linear production, the input to the factory is not a stream of nearly identical new parts. Each returned part carries an individual history of usage intensity; damage pattern, residual functionality, contamination, and material aging may therefore differ between parts. These differences create distinct condition regimes with continuous variation within each regime. The relevant advantage of a GMM in this setting is that it can retain this structure rather than collapsing all returned parts into one homogeneous population [14,1]. This application specific motivation does not imply that a GMM is universally adequate for every non-Gaussian distribution.

The Collaborative Research Centre 1574 Circular Factory studies the integration of linear and circular production through product–production co-design and the modelling of individual product instances in autonomous, modular production systems [3,13]. In this context, the physical state of a returned product determines whether reuse or further processing is technically feasible. Incomplete knowledge of that state affects the reliability of reuse, inspection, routing, and final quality decisions.

A recent GMM-based uncertainty framework argues that uncertain quantitative attributes in measurement and circular factory systems can be represented as GMMs, propagated by closed-form operations when possible, and reduced to a conventional Gaussian summary whenever only mean and variance are required [6]. The present paper does not repeat that framework. Instead, it focuses on the practical case where an uncertainty must be propagated through a nonlinear engineering model, a decision rule, or a surrogate model for which closed-form propagation is no longer available.

The resulting practical question is simple. Given a GMM and a scalar quantity of interest, such as processing time, quality loss, defect score, or threshold violation probability, how many Monte Carlo samples are enough? In many engineering workflows this number is still selected by habit or by a fixed budget. Too few samples can create false confidence in a decision, whereas too many samples waste computation time and make uncertainty propagation unattractive in iterative planning loops.

The goal of this paper is therefore to describe a practical uncertainty propagation workflow for heterogeneous return streams. A GMM is used to retain the multimodal structure of the input uncertainty, while an automatic sampling layer propagates this structure through nonlinear process or quality models. The user specifies a tolerance and a confidence level; the sampler generates GMM samples, monitors the estimated numerical error of the result, and stops when the requested precision is reached or when the sampling budget is exhausted. Accordingly, the contribution is the engineering integration of this established sampling principle into a GMM-based workflow for nonlinear circular factory models, rather than a new fixed-width confidence interval theory.

2 Gaussian mixtures as uncertainty models in circular factories

An uncertain quantity $X \in \mathbb{R}^d$ is represented by the following parameter tuple and density

$$g_X = (K_X, \{\pi_{Xi}, m_{Xi}, \Sigma_{Xi}\}_{i=1}^{K_X}), \qquad G_X(x|g_X) \quad = \sum_{i=1}^{K_X} \pi_{Xi} N_X(x|m_{Xi}, \Sigma_{Xi}) \tag{1}$$

Here, K_X is the number of regimes, π_{Xi} is the weight of regime i, and m_{Xi} and Σ_{Xi} describe the local mean and covariance inside that regime. In a circular factory, such regimes may correspond to different returned part conditions, suppliers, repair histories, contamination states, or material classes.

Closed-form GMM operations are treated in detail in [6]. Here, the GMM representation is used as input to an automatic Monte Carlo sampling layer for cases where the next propagation step is nonlinear or algorithmic.

For scalar reporting, the first two moments are often the most familiar summary for engineering communication, because they correspond to the conventional reporting format of a nominal value together with an uncertainty measure [10,6]. They do not replace the full GMM representation, but provide a Gaussian fallback when a compact report is required. The mean and covariance of a GMM are available in closed form:

$$\mathbb{E}[X] = \sum_{i=1}^{K_X} \pi_{Xi} m_{Xi}, \qquad \mathrm{Cov}[X] = \sum_{i=1}^{K_X} \pi_{Xi}(\Sigma_{Xi} + m_{Xi} m_{Xi}^\top) - \mathbb{E}[X]\mathbb{E}[X]^\top \tag{2}$$

These formulas are useful for validation and for reporting. However, they do not solve the main problem of this paper, because the quantity of interest is often not X itself but

$$Y = g(X) \tag{3}$$

where g may be a nonlinear engineering model, a process rule, an inspection policy, or a learned surrogate.

3 When Monte Carlo is needed

For several GMM operations, closed-form propagation is available and should be preferred. However, many factory quantities are not produced by algebraic GMM operations, but by nonlinear or algorithmic modules, for example routing decisions, threshold-based quality rules, nonlinear fatigue relations, or processing time models.

In these cases, Monte Carlo propagation is a practical fallback. Samples are drawn from the input GMM, evaluated by the model g, and the resulting values $Y_1, \ldots, Y_N$ are used to estimate the required scalar quantity. The general role of Monte Carlo propagation is discussed in [10,12,5]. Here, we focus on the main stopping question. How many propagated samples are sufficient for a prescribed numerical precision?

4 Monte Carlo estimators and confidence interval based stopping

Sampling from the GMM input is performed by first drawing a component index according to the mixture weights and then drawing from the corresponding Gaussian component. This preserves the mixture weights in expectation and produces samples from the full multimodal input distribution, not from a Gaussian approximation of its first two moments [1].

Let $Y = g(X)$ denote a scalar quantity of interest obtained from the uncertain input X. Given samples $Y_1, \ldots, Y_N$, the raw moment estimator and the mean stopping rule are

$$\begin{aligned} \hat{M}_{k,N} &= \frac{1}{N}\sum_{j=1}^{N} Y_j^k, \qquad \hat{m}_N = \hat{M}_{1,N}, \\ w_m(N) &\approx z_{1-\alpha/2}\frac{\hat{s}_N}{\sqrt{N}}, \qquad w_m(N) \le \varepsilon_m. \end{aligned} \tag{4}$$

Here, $\hat{s}_N$ denotes the empirical standard deviation of $Y_1, \ldots, Y_N$, and ε_m is the user specified tolerance. The parameter α specifies the confidence level $1-\alpha$, and $z_{1-\alpha/2}$ is the corresponding standard normal quantile. For example, a nominal 95% level uses $\alpha = 0.05$ and $z_{0.975} \approx 1.96$.

In engineering terms, $w_m(N)$ estimates the remaining Monte Carlo sampling error in the estimated mean. For example, if $\varepsilon_m = 5\,\mathrm{MPa}$, the algorithm continues sampling until the estimated half width for the mean is at most 5 MPa. This bound concerns numerical sampling precision but it does not represent the physical spread of fatigue strength across components. This follows standard fixed-width and convergence monitoring ideas for Monte Carlo estimation [2,15,9]. The normal approximation is appropriate for the large sample sizes used in the stopping procedure and should be interpreted as an asymptotic precision diagnostic rather than as an exact finite sample interval.

For the second raw moment, each propagated result Y_j is simply replaced by Y_j^2. More generally, the raw moment of order k, $M_k = \mathbb{E}[Y^k]$, is the mean of the transformed random variable Y^k, so the same confidence interval logic can be applied to $Y_1^k, \ldots, Y_N^k$. In the experiment, the second raw moment is monitored with a relative half width based on the empirical standard deviation of the transformed samples Y^2 [15].

For the present circular factory use case, we mainly focus on the mean and the second raw moment, because these provide the variance through

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2. \tag{5}$$

Besides moments, many circular factory decisions require probabilities of crossing a quality or safety threshold. For this purpose, each propagated value is reduced to a binary statement. The indicator is one if the threshold is violated and zero otherwise. The estimated probability is therefore simply the fraction

of propagated samples that violate the threshold. The same confidence interval logic therefore also applies to threshold probabilities. For a threshold y_{crit}, define

$$I_j = \mathbf{1}\{Y_j < y_{\text{crit}}\}, \qquad \hat{p}_N = \frac{1}{N}\sum_{j=1}^{N} I_j,$$
$$w_p(N) \approx z_{1-\alpha/2}\sqrt{\frac{\hat{p}_N(1-\hat{p}_N)}{N}}. \tag{6}$$

Sampling stops once $w_p(N)$ is below the user specified probability tolerance [15].

5 Automatic sampling module for circular factory applications

The proposed sampler is intended as a small software layer between a GMM uncertainty representation and a model or decision module. The user specifies the quantity of interest, the tolerance, the confidence level, the batch size, and the maximum sampling budget. The final sample size is not chosen manually, but selected by the stopping rule.

The module proceeds batch by batch. In each batch, it draws GMM samples by the composition method, evaluates them with the chosen model or decision rule, and updates the current estimate and half width from Section 4. Sampling stops when the requested tolerance is reached. If the tolerance is not reached before the maximum budget, the module returns a warning together with the best available estimate.

In the proof of concept, the batch-wise scheme is applied to three scalar quantities, the mean of Y, the second raw moment estimated from Y^2, and the low strength probability. The variance is then computed from the first two raw moments using Eq. (5). Instead of reporting only a Monte Carlo estimate, the module reports the estimate, the final sample size, the achieved half width, and whether the maximum sampling budget was reached.

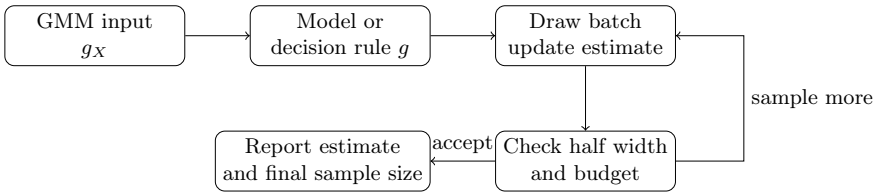


Fig. 1. Automatic GMM sampling loop. The module receives a GMM, a scalar model output, and user precision requirements. It draws samples in batches and stops when the estimated half width satisfies the requested tolerance or when the maximum budget is reached.

6 Circular factory use case

A representative use case is uncertainty propagation in the fatigue strength assessment of returned or treated metallic components. This example matches the Circular Factory setting because the future usability of a component depends not only on its nominal material class, but also on uncertain local material properties after use, inspection, repair, or surface treatment [13,3].

The physical reference is the concept of local fatigue strength for martensitic surface layers and laser hardening layers [16]. In this concept, local fatigue strength is computed from material quantities such as hardness, residual stress, surface roughness, and surface influence. In the proof of concept, the relation is evaluated at one representative critical surface location; a full fatigue strength profile over depth is not modelled.

The material model first derives three intermediate quantities from Vickers hardness HV and the surface roughness parameter R_z as

$$R_m = 3.3\,HV, \qquad R_w^0 = \frac{1.98HV - 0.0011HV^2}{1 + 20.7/HV}, \qquad Y_R(R_z) = 1.49 - 0.471(R_z + 1)^{0.1}. \tag{7}$$

Here, R_m is the static tensile strength, R_w^0 is the local stress free alternating load resistance used as the baseline fatigue strength, and Y_R describes the influence of surface roughness. The symbols follow established materials engineering notation. While R_m and R_w denote strength quantities, R_z is the conventional symbol for the roughness parameter, so the common letter R does not imply the same physical role. The variable z denotes depth below the surface. At the surface, the roughness influence enters through $\omega(z = 0) = Y_R(R_z)$, whereas below the surface $\omega(z > 0) = 1$. Since only the surface point is evaluated here, we use $\omega = Y_R(R_z)$. The gradient support factor $n(z)$ is set to one. The fatigue strength is then evaluated by

$$R_w = \omega R_w^0 n(z) \left(1 - \frac{\sigma_{\mathrm{ES}}}{R_m}\right). \tag{8}$$

Here, σ_{ES} denotes the residual stress and R_w is the resulting local fatigue strength. The stress and strength quantities are reported in MPa. For the sampling experiment, the uncertain input and scalar output are

$$X = (HV, \sigma_{\mathrm{ES}}, R_z), \qquad Y = R_w = g(X). \tag{9}$$

A GMM can describe different coupled surface condition regimes, for example wear affected, laser hardened, or damaged states. Each sample of X is passed through Eqs. (7) and (8). If an applied local stress amplitude $\sigma_{\mathrm{a,loc}}$ is available, the output can alternatively be written as a safety margin,

$$M = R_w - \sigma_{\mathrm{a,loc}}. \tag{10}$$

The proposed sampling module draws samples from the GMM representation of X, evaluates the fatigue model in Eq. (8) for each sample, and estimates quantities such as the mean fatigue strength, the second raw moment used to compute variance, or the probability of a low strength event

$$p_{\text{low}} = \mathbb{P}(R_w < R_{\text{crit}}) \tag{11}$$

where R_{crit} is an illustrative threshold chosen for the decision task. The automatic stopping rule determines how many samples are needed until the requested tolerance for the chosen output is reached.

The experiment does not claim a new fatigue model. It uses an existing fatigue assessment relation as a nonlinear engineering model and tests whether the required Monte Carlo sample size can be controlled automatically for multimodal uncertain inputs.

7 Proof of concept experiment

The proof of concept was implemented in Python using the `gmtorch` package, which provides a PyTorch-based implementation of Gaussian mixture models and sampling operations [4].

The numerical scenario uses a three component GMM over hardness, residual stress, and surface roughness. The components represent a wear affected or untreated regime, a laser hardened regime, and a rare damaged regime. Depending on the chosen parameters, the rare damaged regime can contribute to the low strength part of the propagated output because it combines lower hardness, higher roughness, and less favorable residual stress. The experiment therefore tests the sampling layer on the nonlinear mapping defined by Eq. (8).

The domain check accepts only samples with $250 < HV < 1000$, $-1200 < \sigma_{\text{ES}} < 500$, and $1 < R_z < 25$. These bounds define the physically admissible input range used in the numerical experiment. The GMM parameters are synthetic and are not experimentally identified material parameters.

The propagated output of Eq. (8) is not well described by a single symmetric density; a low strength mode remains visible after nonlinear propagation. To test what is lost by an early Gaussian approximation, the same input uncertainty was also replaced by a single Gaussian constructed from the first two input moments of the three component GMM. Samples outside the physical input domain were rejected before evaluating Eq. (8). The same domain check was applied to both GMM and Gaussian samples. After this rejection step, the accepted Gaussian sample is no longer exactly moment matched. The comparison is therefore used as a diagnostic for early Gaussian approximation under physical input constraints.

The comparison in Fig. 2 shows why the mixture structure should be preserved during propagation. Replacing the returned part population by a single Gaussian can change the estimated probability of violating the quality threshold. The GMM input preserves the regime structure of the propagated distribution, while the Gaussian approximation produces a smoother output distribution.

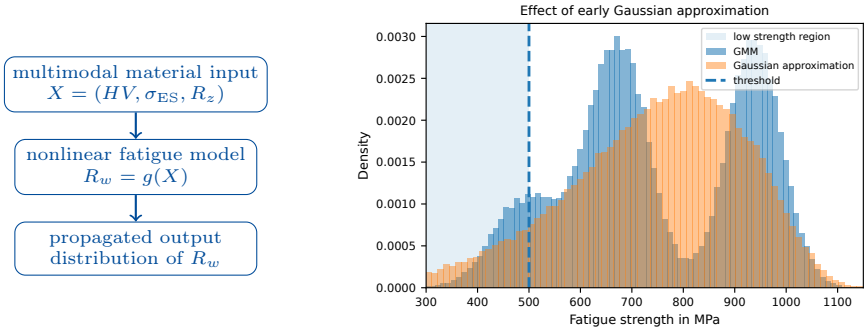


Fig. 2. Nonlinear propagation of the multimodal material input through g. The schematic shows the mapping from $X = (HV, \sigma_{ES}, R_z)$ to $R_w = g(X)$, while the plot compares the propagated output distributions for the GMM input and its Gaussian approximation. Both inputs are checked against the physical domain before evaluation.

The shaded region marks the illustrative low strength event $R_w < R_{crit}$ with $R_{crit} = 500\,\text{MPa}$. This region defines the event used in Eq. (11), but it is not introduced as a validated fatigue design limit.

In Fig. 3, the estimate stabilizes close to 0.09, that is, about nine percent of the propagated samples belong to the low strength region. The shaded band is the approximate 95 percent interval obtained from the Bernoulli half width. The sampling procedure stops when the interval half width falls below the requested probability tolerance.

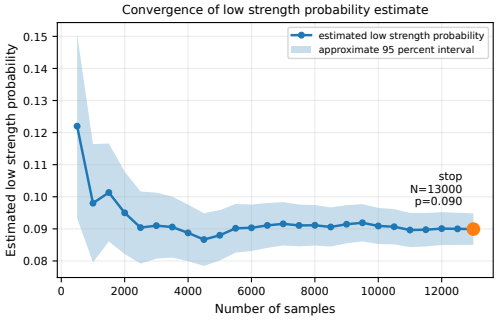


Fig. 3. Convergence of the estimated low strength probability. The shaded band shows the approximate 95 percent interval, and the marked point indicates where the requested probability tolerance is reached.

The automatic stopping behavior for the first two raw moments is reported in Fig. 4. For the mean fatigue strength, the tolerance is expressed in MPa. The selected sample size increases from $N = 500$ for a tolerance of 20 MPa to

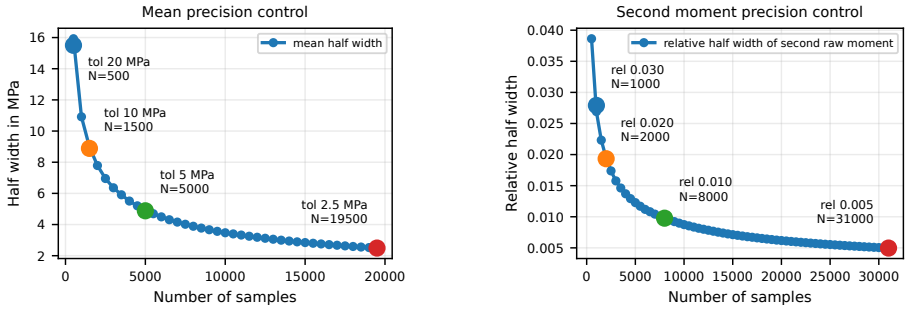


Fig. 4. Automatic precision control for the estimated mean fatigue strength and the second raw moment. The first plot shows the half width in MPa, while the second shows the relative half width obtained from R_w^2.

$N = 19500$ for a tolerance of 2.5 MPa. For the second raw moment, the same fixed-width idea is applied to the transformed samples R_w^2. Since the plotted quantity is a relative half width, the vertical axis is dimensionless, while the transformed samples have unit MPa^2. The selected sample size increases from $N = 1000$ for a relative tolerance of 0.030 to $N = 31000$ for a relative tolerance of 0.005.

8 Conclusion

The central message of this paper is that the uncertainty representation should not be simplified before model evaluation merely because the later computation requires samples. For circular factory applications, this matters because returned parts may belong to different material or process regimes. A GMM can express such regimes explicitly, while the proposed sampling module keeps this structure available during nonlinear propagation.

The proposed module acts as a controlled propagation layer between GMM-based uncertainty descriptions and scalar engineering quantities of interest. It preserves the multimodal input during nonlinear model evaluation and reports scalar quantities together with achieved numerical precision and required sample size. The fixed-width stopping rule makes the numerical effort visible and turns the final sample size into an output of the analysis. The contribution therefore lies in this engineering integration rather than in introducing a new Monte Carlo convergence theory.

The present example remains synthetic and local to one representative surface point. Future work should identify the GMM parameters from measurement data, connect the module to richer process chains, and study how precision controlled uncertainty outputs can be used in closed loop decision support.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
2. Chow, Y.S., Robbins, H.: On the asymptotic theory of fixed-width sequential confidence intervals for the mean. The Annals of Mathematical Statistics **36**(2), 457–462 (1965). https://doi.org/10.1214/aoms/1177700156
3. Collaborative Research Centre 1574: Circular Factory for the perpetual product. Karlsruhe Institute of Technology, project information pages (2026), accessed 2026
4. Darijani, A.: gmtorch: Gaussian mixture models in PyTorch. Python package, PyPI (2025), version 0.0.1. Available at https://pypi.org/project/gmtorch/
5. Darijani, A.: Uncertainty quantification and the need for better understanding of Monte Carlo and random number generation. In: Proceedings of the 2024 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. Karlsruher Schriften zur Anthropomatik, vol. 70. KIT Scientific Publishing, Karlsruhe (2025)
6. Darijani, A., Beyerer, J., Hajseyed Nasrollah, Z.S., Hoffmann, L., Heizmann, M.: Comprehensive description of uncertainty in measurement for representation and propagation with scalable precision (2026). https://doi.org/10.48550/ARXIV.2603.20365
7. Guide, V.D.R.: Production planning and control for remanufacturing: Industry practice and research needs. Journal of Operations Management **18**(4), 467–483 (2000). https://doi.org/10.1016/S0272-6963(00)00034-6
8. Ilgin, M.A., Gupta, S.M.: Environmentally conscious manufacturing and product recovery: A review of the state of the art. Journal of Environmental Management **91**(3), 563–591 (2010). https://doi.org/10.1016/j.jenvman.2009.09.037
9. Janssen, H.: Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. Reliability Engineering & System Safety **109**, 123–132 (2013). https://doi.org/10.1016/j.ress.2012.08.003
10. Joint Committee for Guides in Metrology: Evaluation of measurement data – supplement 1 to the guide to the expression of uncertainty in measurement – propagation of distributions using a Monte Carlo method. Tech. Rep. JCGM 101:2008, JCGM (2008). https://doi.org/10.59161/JCGM101-2008
11. Kirchherr, J., Reike, D., Hekkert, M.: Conceptualizing the circular economy: An analysis of 114 definitions. Resources, Conservation and Recycling **127**, 221–232 (2017). https://doi.org/10.1016/j.resconrec.2017.09.005
12. Kroese, D.P., Taimre, T., Botev, Z.I.: Handbook of Monte Carlo Methods. Wiley (2011). https://doi.org/10.1002/9781118014967
13. Lanza, G., Deml, B., Matthiesen, S., Martin, M., Brützel, O., Hörsting, R.: The vision of the Circular Factory for the perpetual innovative product. at – Automatisierungstechnik **72**(9), 774–788 (2024). https://doi.org/10.1515/auto-2024-0012
14. McLachlan, G.J., Peel, D.: Finite Mixture Models. Wiley (2000). https://doi.org/10.1002/0471721182
15. Serfling, R.J.: Approximation Theorems of Mathematical Statistics. Wiley (1980). https://doi.org/10.1002/9780470316481
16. Winderlich, B.: Das konzept der lokalen dauerfestigkeit und seine anwendung auf martensitische randschichten, insbesondere laserhärtungsschichten. Materialwissenschaft und Werkstofftechnik **21**(10), 378–389 (1990). https://doi.org/10.1002/mawe.19900211006