Advanced Energy Materials

ADVANCED ENERGY MATERIALS
www.advenergymat.de

RESEARCH ARTICLE OPEN ACCESS

Model-Based Bayesian Optimization for Organic Photovoltaics: Combining Bayesian Optimization With Physical Domain Knowledge

Leonard Christen[1] | Thomas Kirchartz[1,2]

[1]IMD3-Photovoltaik, Forschungszentrum Jülich, Jülich, Germany | [2]Faculty of Engineering and CENIDE, University of Duisburg-Essen, Duisburg, Germany

Correspondence: Leonard Christen (l.christen@fz-juelich.de) | Thomas Kirchartz (t.kirchartz@fz-juelich.de)

Received: 16 April 2026 | **Revised:** 2 June 2026 | **Accepted:** 18 June 2026

Keywords: bayesian process optimization | machine learning | model-based process optimization | organic photovoltaics

ABSTRACT

Efficient process optimization is of crucial importance in materials science and industrial manufacturing. Bayesian optimization has proven to be an effective method for optimizing time-consuming experiments; however, it is typically carried out as a black-box approach without utilizing prior physical knowledge. In this work, we apply a model-based Bayesian optimization approach in which a physical model of the solar cell is integrated into the optimization process. Simultaneous exploration and exploitation of the parameter space is enabled by the acquisition function, as is customary for Bayesian Optimization. We show that the so-called Knowledge Gradient fits particularly well for the model-based approach. The method was validated both experimentally, using the PTQ10:BTP-eC9 material system, and statistically, using a benchmark function specifically developed for organic photovoltaics. The results show that the model-based approach is superior to the conventional black-box approach and that the global optimum can be identified more reliably. This work demonstrates the potential of integrating domain knowledge into Bayesian optimization algorithms and opens up new perspectives for accelerated process optimization in photovoltaic research.

1 | Introduction

Machine-learning-based methods for process optimization have become increasingly important in materials science and industrial manufacturing in recent years, including in the field of photovoltaics [1–7]. Bayesian Optimization has emerged as a particularly effective methodology for process optimization within the field of materials science. This is primarily due to its suitability for optimization problems where sampling from the parameter space is time-intensive, which is commonly encountered in scenarios in laboratory settings or industrial manufacturing processes [8]. Bayesian Optimization generally operates with little or no explicit physical knowledge of the system being optimized. The nature of this black-box approach implies that measurements unrelated to the figure of merit cannot directly influence the optimization trajectory. Additional measurements are therefore usually employed for retrospective analysis rather than for accelerating the optimization process, which questions the overall relevance of sophisticated characterization and human-level understanding for the purpose of the optimization process.

Here, we discuss an enhanced approach that integrates such knowledge in the form of physical models and characterization data into Bayesian Optimization applied to finding the best processing parameters for a given type of organic solar cells. Unlike silicon-based photovoltaics, organic photovoltaics relies on molecular semiconductors for light absorption and charge-carrier transport, enabling the production of flexible, semi-transparent, and lightweight solar modules [9]. These

properties of organic solar cells enable specialized applications; the integration of photovoltaics into buildings or windows is just one of many examples [10]. One of the key advantages of molecular semiconductors is the huge chemical flexibility that allows chemists to synthesize many different molecules that can be designed or identified to have specific desired properties [11, 12]. However, this increases the search space for identifying ideal molecules, and it also increases the difficulty of finding optimum processing conditions, as those have to be optimized separately for each new set of molecules to be tested for a specific application such as photovoltaics [13–15]. Thus, organic photovoltaics often appears like a two-needles-in-two-haystacks problem, where the first haystack is the multitude of different conceivable molecules and the second haystack is the multidimensional process parameter space; furthermore, both are correlated.

Thanks to promising research, new efficiency records are regularly being set for organic solar cells, and the 20% efficiency mark has already been overcome [16–18]. However, the potential of organic photovoltaics has not yet been fully exploited. Furthermore, in addition to efficiency, it is also important to maximize stability and recyclability, optimize flexibility, and minimize production costs, all areas in which organic photovoltaics shows great potential [19–23]. An efficient and effective process optimization is therefore essential here, and the use of domain knowledge offers a way to improve (Bayesian) process optimization.

2 | Model-Based Bayesian Optimization

While classical approaches to process optimization, such as the one-factor-at-a-time method, are straightforward to implement in the laboratory, they quickly reach their limits when dealing with higher-dimensional parameter spaces containing correlated parameters [24]. Modern and more sophisticated process optimization is increasingly relying on machine learning-based approaches [3, 6]. If physical domain knowledge remains unused in this context and the optimization algorithm relies exclusively on a non-physical, mathematical description, this is referred to (below) as black-box optimization [25]. Here, the figure of merit is modelled as a function of the process parameters contained in the vector θ_{pro} (such as donor-acceptor ratio, solvent concentration, spin speed, annealing conditions, and many more). Conversely, should physical domain knowledge be employed during the optimization, we will refer to this as model-based optimization. It should be noted that the idea of using additional knowledge about the system for optimization purposes is referred to differently depending on the context. Examples include terms such as 'gray-box' [26], 'composite functions' [27], or 'physics-informed' [28, 29]. However, in this model-based optimization approach, θ_{pro} is associated with model parameters θ_{mod}, and the figure of merit is calculated using a corresponding physical model. In photovoltaics, the figure of merit is often the solar cell efficiency η [30]. However, alternative metrics such as stability, cost, recyclability, or mechanical flexibility may also be incorporated depending on the application context [23, 31, 32]. A methodology to combine such alternatives with the solar efficiency has been reported [33]. For further clarification, please refer to Figure 1. In the following, we will take a closer look at the requirements for the model parameters and the model itself, to then examine the actual process optimization in more detail later.

2.1 | The Physical Model

In principle, a wide range of physical models could be adapted for model-based process optimization. The decisive requirement for the physical model is the establishment of a mathematical link between measurable physical quantities (namely, the model parameters θ_{mod}) and the final figure of merit (as mentioned, in photovoltaics, this, for example, might be the solar cell efficiency η). Given that a Bayesian optimization framework requires frequent model evaluations, meaning that the figure of merit needs to be calculated for different parameters very often, a preference should be given to analytical formulations over computationally intensive numerical approaches. In principle, however, numerical models are also conceivable. Ideally, the model should be as simple as possible while incorporating all relevant physical domain knowledge that is necessary for its purpose.

In the Present Proposal for a Possible Model, the Solar Cell Efficiency Is Calculated From the JV Curve via

$$\eta = \frac{P_{max}}{P_{sun}} = \frac{J_{mpp} V_{mpp}}{P_{sun}} \tag{1}$$

where P_{max} is the maximum power density at the maximum power point at the current density $J = J_{mpp}$ and voltage $V = V_{mpp}$. The efficiency calculates the ratio of P_{max} to the irradiance P_{sun}. The typically used equation for the current-voltage curve of a solar cell is

$$J\ (V) = J_0 \cdot \left[\exp\left(\frac{V}{n_{id} V_{th}}\right) - 1\right] - J_{illum} \tag{2}$$

with the dark saturation current density J_0, the ideality factor n_{id}, the thermal voltage $V_{th} = k_B T/q$, and the current density generated by light illumination J_{illum}. In organic solar cells, voltage-dependent charge-collection losses lead to the situation that Equation (2) has to be somewhat modified to accommodate additional losses that primarily reduce the fill factor ($FF = J_{mpp} V_{mpp}/(J_{SC} V_{OC})$). An approach suggested by Neher et al. involves replacing the ideality factor n_{id} with an apparent ideality factor n_{app} that includes the effect of charge-collection losses, especially close to open circuit [34]. There are situations where it is useful to additionally include the dark shunt resistance R_p, leading to an approximately linear behavior for dJ/dV around $V = 0$. This modification is particularly relevant for low-efficiency solar cells with pinholes or for measurements at low light intensities relevant for indoor applications. In both cases, the dark shunt may contribute so strongly to the current-voltage curve that the parameter n_{app} cannot capture the behavior anymore. With these modifications, we arrive at

$$J\ (V) = J_0 \cdot \left[\exp\left(\frac{V}{n_{app} V_{th}}\right) - 1\right] - J_{illum} + \frac{V}{R_p} \tag{3}$$

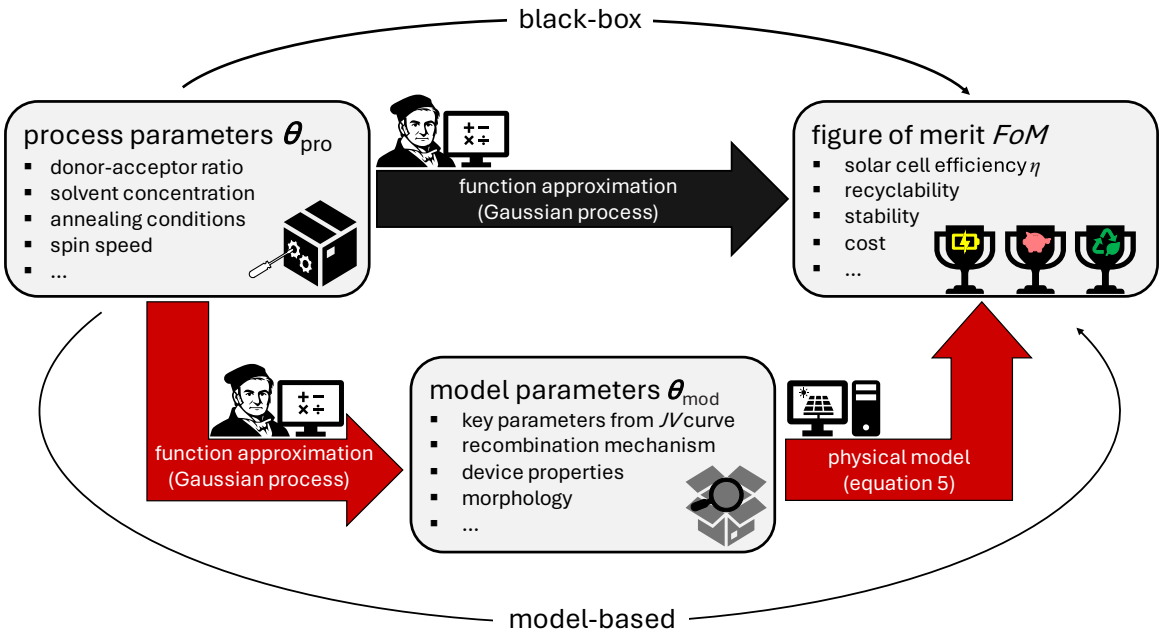

FIGURE 1 | Comparison of black-box and model-based approaches. The top route illustrates the black-box approach, where the figure of merit is directly associated with the process parameters θ_{pro}. The lower route illustrates the model-based approach. Here, additional model parameters θ_{mod} are associated with the process parameters and then used to calculate the figure of merit using a physical model.

Using $J\,(V = V_{OC}) = 0$ Leads to

$$J_0 = \left(\eta_{eh} J_{gen} - \frac{V_{OC}}{R_p}\right) \cdot \exp\left(\frac{-V_{OC}}{n_{app} V_{th}}\right) \tag{4}$$

Here, J_{illum} is replaced by the product of the total generation current density J_{gen} and the charge-carrier dissociation efficiency η_{eh}. Reinserting J_0 into the *JV* curve then gives us the final model, which we propose here

$$J\,(V, \theta_{mod}) = \left(\eta_{eh} J_{gen}\,(d) - \frac{V_{OC}}{R_p}\right) \cdot \exp\left(\frac{V - V_{OC}}{n_{app} V_{th}}\right) - \eta_{eh} J_{gen}\,(d) + \frac{V}{R_p} \tag{5}$$

with the final physical model parameters $\theta_{mod} = (V_{OC}d, \eta_{eh}, n_{app}, R_p)$. For more information regarding the model and a more detailed derivation, please refer to the Supporting Information. In this representation for the physical model, $J_{gen}(d)$ depends completely on the active layer thickness and the optical properties (assumed to be measured and constant as a function of process parameters) and is therefore not listed as a separate model parameter. The isolated influence of a change in the individual model parameters is illustrated in Figure 2. We emphasize that the model parameters are usually correlated, and in real laboratory conditions, it is difficult to isolate the change of a single model parameter. An overview of the individual model parameters can be found in Table 1.

We first examine the active layer thickness d in more detail. Conventionally, an organic solar cell comprises a multilayer stack wherein a transparent conductive substrate (such as ITO-coated glass) serves as the foundation. On this substrate, a layered architecture is deposited, consisting of a photoactive layer sandwiched between an electron transport layer and a hole transport layer, followed by a metallic contact. Due to the highly reflective nature of the metal electrode and the characteristically thin layers used in organic photovoltaics, optical interferences arise within the device. These phenomena are strongly dependent on individual layer thicknesses. To quantify these interference effects, the optical properties (refractive index n and extinction coefficient k) of the constituent layers are determined experimentally. Subsequently, interference patterns can be computationally modelled using the transfer matrix method to simulate their behavior [35]. In the presented case, we keep all layer thicknesses constant during process optimization, with the exception of the active layer thickness, which serves as a primary parameter for device optimization. For a given layer stack, the charge-carrier absorption can then be calculated depending on the layer thickness of the active layer. Using the elementary charge q, this then directly results in $J_{gen}(d)$. It should be noted that this restriction is not fixed, and the model can be readily adapted, albeit at a higher computational cost, to account for simultaneous variations in multiple layer dimensions. In this framework, the active layer thickness is experimentally characterized using capacitance measurements at reverse bias in the dark, where the active layer is fully depleted, and the capacitance is dominated by the geometrical capacitance $C = \varepsilon_0 \varepsilon_r A/d$. Here, the solar cell area is denoted by A, and ε_0 denotes the vacuum permittivity. This non-destructive method of inferring the thickness d is complemented with profilometer measurements on a subset of samples to determine the relative permittivity ε_r and monitor potential systematic errors in the measurement using the reverse-bias capacitance.

The current density $J_{gen}(d)$ due to photogeneration at short circuit is derived from the active layer thickness d and represents an idealized value that ignores any exciton dissociation or charge

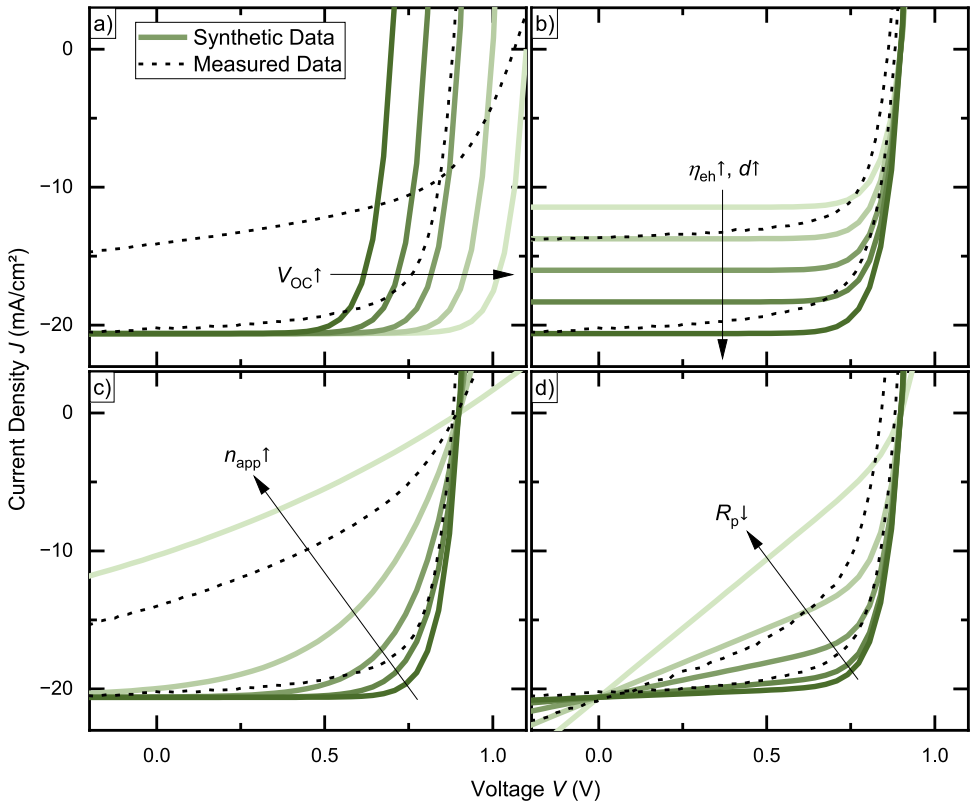

FIGURE 2 | Various *JV* curves are shown in green for the isolated variation of the individual model parameters, namely (a) the open-circuit voltage V_{OC}, (b) the active layer thickness d as well as the charge-carrier dissociation efficiency η_{eh},(c) the apparent ideality factor n_{app}, and (d) the shunt resistance R_p. The dashed lines show some measured *JV* curves. The selection is intended to demonstrate the influence of the model parameters on the *JV* curves in a non-synthetic scenario. Here, the parameters are typically correlated.

TABLE 1 | Parameters of the physical model for the *JV* curve as shown in Equation (5).

Model parameter		Description	Measurement
Open circuit voltage	V_{OC}	$J(V_{OC}) = 0$	illuminated *JV* curve
Active layer thickness	d	thickness of the light-absorbing material, directly leads to $J_{gen}(d)$	capacitance measurements (from geometrical capacitance at reverse bias) or profilometer
Charge-carrier dissociation efficiency	η_{eh}	incorporates optical losses and exciton dissociation losses	illuminated *JV* curve, inferred with Equation (6)
The apparent ideality factor	n_{app}	charge-carrier losses close to open circuit	illuminated *JV* curve, inferred with Equation (5)
The dark shunt resistance	R_p	$R_p = \frac{dV}{dJ_{dark}}\|_{V=0}$	dark *JV* curve

carrier recombination losses within the active layers. The exciton dissociation losses are caused by excitons recombining before they are dissociated into electron-hole pairs and can be taken into account using the charge-carrier dissociation efficiency η_{eh}. Furthermore, η_{eh} also considering deviations between the optical model and the actual rate of photon absorption, as this deviation is impossible to discriminate from exciton dissociation losses given the typical experimental observables available for a device optimization loop. However, this parameter is not directly measured, but estimated by the following equation using the short-circuit current density J_{SC}, which is obtained from the illuminated *JV* curve, as is the case for the V_{OC}. Using Equation (5) and $J(V = 0) = -J_{SC}$ leads to (see Supporting Information)

$$\eta_{eh} = \frac{J_{SC} - \frac{V_{OC}}{R_p}\exp\left(\frac{-V_{OC}}{n_{app}V_{th}}\right)}{J_{gen}(d)\left(1 - \exp\left(\frac{-V_{OC}}{n_{app}V_{th}}\right)\right)} \quad (6)$$

The objective of the apparent ideality factor n_{app} is to adequately capture charge-collection losses in the solar cell, in particular in situations where those losses appear mostly close to open circuit and where they do not lead to S-shaped current-voltage curves.

As mentioned above, the concept has been introduced by Neher et al. [34]. and further used and analyzed, for instance, by Saladina et al. [36, 37]. The apparent ideality factor can be measured using light intensity-dependent *JV* curves. In this study, the measured *JV* curve is compared with the model $J(V, \theta_{mod})$ to infer the apparent ideality factor.

The inclusion of the dark shunt resistance R_p is not universally required. Depending on the parameter space and experimental conditions, the parameter can be useful for our model in cases where the measured *JV* characteristics cannot be accurately reproduced by the model using the above-described apparent ideality factor (n_{app}) alone. However, if the process optimization is constrained to devices exhibiting high shunt resistances, this parameter can be neglected, thereby simplifying the model. Under such conditions, the terms V_{OC}/R_p and V/R_p approach zero, and their influence becomes negligible, leading to

$$J\ (V, \theta_{mod}) = \eta_{eh}\ J_{gen}\ (d) \left(\exp\left(\frac{V - V_{OC}}{n_{app} V_{th}} \right) - 1 \right) \qquad (7)$$

It should be noted that a different physical model can, of course, also lead to different model parameters. These parameters can also include more fundamental material properties, such as charge-carrier mobilities or recombination characteristics, as well as device metrics such as the fill factor. A primary criterion for selecting measurable quantities is their direct or indirect accessibility, ideally with low experimental effort. The additional measurement overhead must always be weighed against the potential time savings enabled by the more efficient model-based process optimization. For further information regarding the measurement methods and data evaluation, please refer to the Supporting Information.

2.2 | Bayesian Process Optimization

Using the presented model and its model parameters, the subsequent process optimization can now be performed. As mentioned above, in a classical black-box optimization approach, the efficiency η is directly correlated with the process parameters, yielding $\eta(\theta_{pro})$. In Bayesian Optimization, however, this estimated $\eta(\theta_{pro})$ is not only a single value, but a probability distribution. This is achieved using a Gaussian process, and it is essential because it enables the optimization process to simultaneously explore and exploit the parameter space effectively. The former actively evaluates uncertain or 'unexplored' areas of the parameter space, while the latter prioritizes high-probability candidates in promising regions. The precise balance between exploration and exploitation is controlled by the so-called acquisition function, which will be examined in more detail below.

In this probabilistic relation between η and θ_{pro}, the proposed physical model must now be incorporated. One approach to achieve this is to compute a separate Gaussian process for each of the model parameters, namely $V_{OC}(\theta_{pro})$, $d(\theta_{pro})$, $\eta_{eh}(\theta_{pro})$, $n_{app}(\theta_{pro})$, and $R_p(\theta_{pro})$, summarized as θ_{mod}. To calculate the probability distribution for the efficiency in the case of a specific set of process parameters, first, thousands of samples are drawn from each model parameter, leading to five different sample sets, one for each model parameter. Combining these samples using the physical model $J(V,\theta_{mod}(\theta_{pro}))$ gives thousands of different *JV* curves for each θ_{pro}, directly leading to the distribution for η. A corresponding illustration can be found in Figure 3. As mentioned above, the model parameters are usually correlated. Training an individual Gaussian process for each model parameter does not inherently account for such correlations. However, we decided to keep the model simple. More sophisticated approaches that inherently model such correlations should be part of future research.

It is worth noting that the resulting distribution function no longer constitutes a Gaussian process. Therefore, some adjustments need to be made when calculating the acquisition function, $a(\theta_{pro},\boldsymbol{\Omega}_{hyp})$. Here, $\boldsymbol{\Omega}_{hyp}$ denotes a vector of hyperparameters that can be adjusted. A straightforward example could be the so-called Upper Confidence Bound, where $a_{UCB}\,(\theta_{pro},\lambda) = \mu(\theta_{pro}) + \lambda\sigma(\theta_{pro})$ applies. Here, the expected value μ is modified by the product of the hyperparameter λ and the standard deviation σ of the Gaussian distribution function. Smaller values for λ lead to a more exploitative acquisition; higher values lead to a more explorative acquisition. However, the idea of the Upper Confidence Bound is not limited to Gaussian distributional settings. In a non-Gaussian setting, instead of the standard deviation σ, for example, the distance between the expected value and the upper 95%-quantile should be modified with the help of λ. Nevertheless, this does not limit the capability for simultaneous exploration and exploitation, which is crucial for Bayesian Optimization.

The choice of acquisition has a decisive influence on Bayesian Optimization. For a brief overview of a few selected acquisition functions, please refer to Table 2. Different acquisition functions are recommended for different situations. In addition to the aforementioned Upper Confidence Bound, frequently used acquisition functions include the Probability of Improvement and the Expected Improvement [38]. These three functions are by far the most commonly used, whether in plain vanilla or sophisticated forms [39–47]. However, we will use and discuss an alternative acquisition function for the proposed model-based Bayesian Optimization approach, namely the Knowledge Gradient, as literature suggests that this performs particularly well for our model-based approach [48, 49]. Since the Knowledge Gradient is a less well-known acquisition function, the following section explains the basic idea behind it and how it can be applied in our case.

While, for example, Expected Improvement asks how likely it is that the next measurement will be better than the previous optimum, the strategy of the Knowledge Gradient uses a one-step lookahead strategy to optimize the entire surrogate model. The following outlines the idea in question. We assume we already have a model-based surrogate model trained with n data points. In our case, this means that n devices were fabricated and characterized in detail for n different process parameter combinations to train individual Gaussian processes for all given model parameters. As mentioned, combining the Gaussian processes with the physical model leads to the final distribution function, the model-based surrogate model $\eta(\theta_{mod}(\theta_{pro}))$. The current optimum η_n^* is now sought within this surrogate

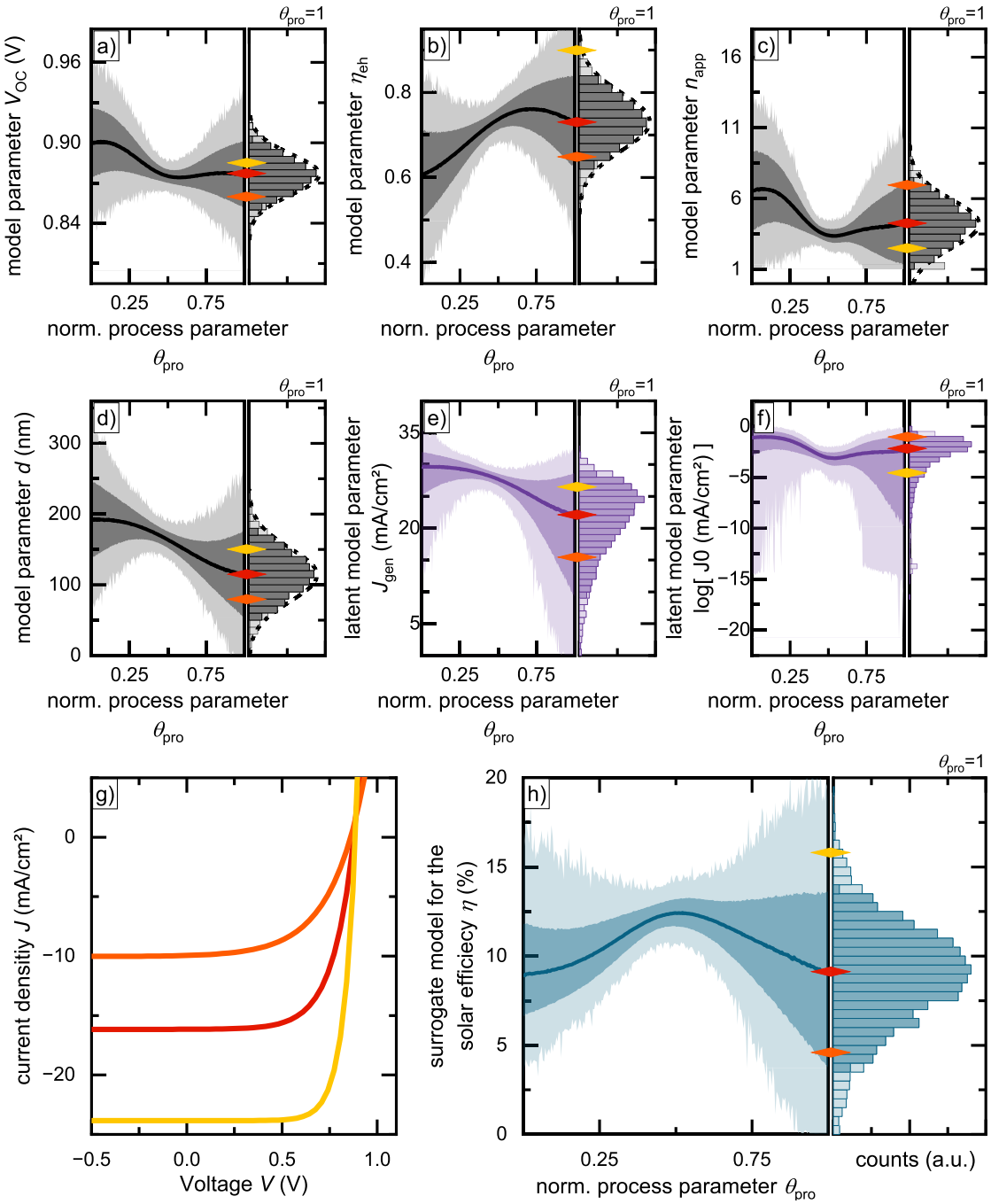

FIGURE 3 | Illustration of how the surrogate model is built. (a—d) For each model parameter, a Gaussian process is trained, as shown on the left side of each graph. The expected value within the 95% confidence interval within the min-max range is shown here for all samples. We neglect the shunt resistance and reduce the process parameter space to one dimension for this illustration. For a given process parameter (e.g., at $\theta_{pro} = 1$), a few thousand data points are sampled from each Gaussian process. The entire distribution is on the right side of each figure. The location of three of those samples is indicated using fire colors. (e,f) Using the physical model, the distribution for latent parameters such as J_{gen} and J_0 can be calculated. Note that, from this point on, the distribution is no longer Gaussian. (g) The *JV* curves for the three example samples. (h) The final surrogate model and the sample distribution at $\theta_{pro} = 1$.

model based on n measuring points. To calculate the Knowledge Gradient for any θ_{pro} in the parameter space, an estimate is made of the influence that a measurement taken at this θ_{pro} would have on the entire surrogate model. For this purpose, a value is drawn from the corresponding distribution at exactly this $\theta_{pro} = \theta_{pro,n+1}$ for each model parameter, and the final surrogate model is recalculated, now based on n+1 measurements, whereby the new measurement is only assumed theoretically. This is often also referred to as fantasizing [51, 52]. The new surrogate's optimum after n+1 measurement points η^*_{n+1} and the optimum after n measurement points η^*_n are compared; in detail, the latter is subtracted from the former. Since the fantasized values are drawn from a distribution function, frequent repetition is recommended. The final value for the knowledge gradient is

TABLE 2 | A selection of different acquisition functions for Bayesian optimization. While UCB, PI, and EI are strategies that are trying to improve the actual optimum within the next round of optimization, the KG strategy is trying to improve future decision-making. Finally, the Entropy Search (ES) is chosen as an example of an information-theoretical approach.

Acquisition function		Description
Upper Confidence Bound [38]	UCB	A straightforward acquisition function that sums the surrogate model's expected value and the uncertainty, multiplied by a weighting factor.
Probability of Improvement [38]	PI	Selects the next point that is most likely to improve upon the best solution found so far.
Expected Improvement [8, 38]	EI	While PI only considers the probability to improve, EI also adds the magnitude of this improvement. The EI always tries to improve over the current best observation, within the very next optimization round.
Knowledge Gradient [8, 48]	KG	Using a one-step lookahead strategy, the KG is trying to improve future decisions, instead of aiming for a direct improvement. Thus, it maximizes the expected increase of the surrogate model (instead of the next measurement, as is for EI).
Entropy Search [8, 50]	ES	An information-theoretic strategy that maximizes the information gain by minimizing the future expected entropy, indicating less uncertainty.

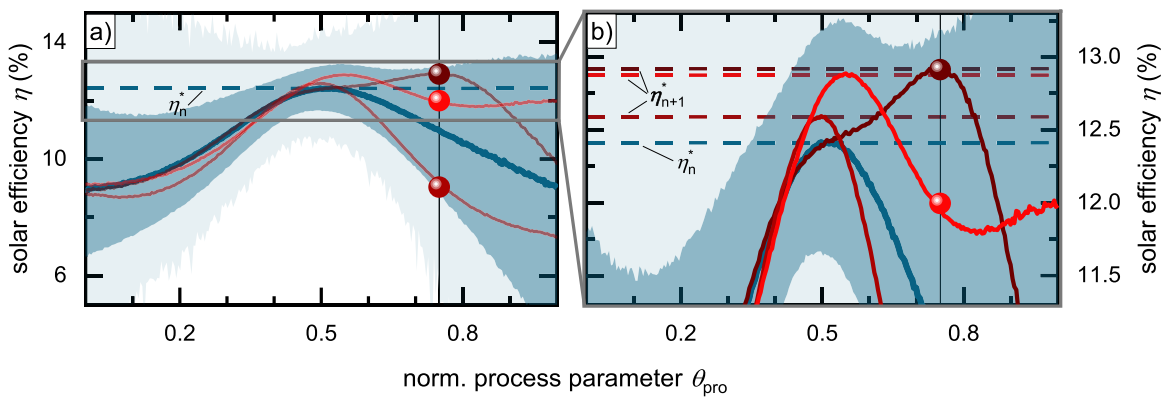

FIGURE 4 | (a) The final surrogate model for the efficiency as a function of a random normalized process parameter θ_{pro}, here in one dimension. After n measurements, the estimated maximum efficiency is given by η_n^*. To calculate the Knowledge Gradient for a position, for example, at $\theta_{pro} = 0.75$, we 'fantasize' the result for an additional measurement n+1. For that, we sample values from the distribution for each θ_{mod} and retrain the Gaussian processes as if we had n+1 measurements. The resulting models for the three examples are shown in red. (b) In the enlarged area, the optima for the fantasized models are indicated using dashed lines. The final Knowledge Gradient for $\theta_{pro} = 0.75$ is calculated by averaging over the distances between η_{n+1}^* and η_n^*.

calculated from the expected value of all draws. Mathematically speaking, the Knowledge Gradient acquisition function is given by

$$a_{KG,n}\left(\theta_{pro}\right) = E\left(\eta_{n+1}^* - \eta_n^* \middle| \theta_{pro,n+1} = \theta_{pro}\right) \tag{8}$$

For an illustration, please refer to Figure 4. This procedure is theoretically repeated for the entire parameter space. The maximum of $a_{KG,n}(\theta_{pro})$ is the best possible measurement point for optimization round n+1 and needs to be executed in the laboratory. For further details regarding its implementation, please refer to the Supporting Information. At this point, however, it should be mentioned that the process described here is computationally expensive.

3 | Experimental Validation

To test the approach described above, we first perform a corresponding process optimization using the PTQ10:BTP-eC9 material system with a device area of 6 mm^2. Figure 5 shows the respective molecules, as well as the used device stack and the energy levels for the active layer from the literature, measured via cyclic voltammetry [53]. For the process optimization, the variable process parameters that constitute θ_{pro} are the donor-to-acceptor ratio $r_{don\text{-}acc}$, the concentration of the solvent o-xylene c_{oxy}, the spin speed ω_{spin}, and the solution temperature during spin coating T_{spin}. For the parameter ranges, we use the limits -1.1 and 1.1 in logarithmic space for $r_{don\text{-}acc}$, which corresponds to ratios of 3:1 and 1:3 for PTQ10:BTP-eC9, c_{oxy} is between 10 and 35 mg/ml; ω_{spin} is between 650 and 5000 rpm, and the limits for

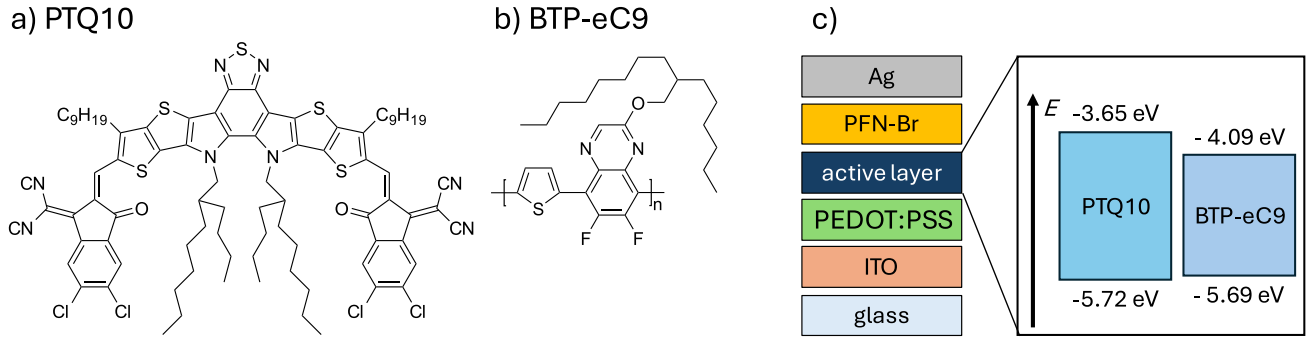

FIGURE 5 | The used molecules for a) PTQ10 and b) BTP-eC9, as well as c) the layer stack and the respective energy levels for the active layer [53].

T_{spin} are 60 and 120°C. All other process parameters remained constant during fabrication. For a detailed description of the device fabrication, please refer to the Supporting Information.

The solar cell efficiency η serves as the figure of merit. The framework conditions for the process optimization were defined as follows. The optimization begins in round 0 with a Latin hypercube sampling for 4 initial proposals. This is followed by 10 optimization rounds, each with two proposals for the next round. Since the Knowledge Gradient does not aim for a direct improvement regarding the next measurement, the last proposal, i.e., the 10th optimization round, is selected differently, by directly choosing the optimum of the surrogate model. This essentially corresponds to an Upper Confidence Bound with $\lambda = 0$. Thus, the final proposal corresponds to an exploitation-only approach. For comparison, the optimization is performed based on the same Latin hypercube sampling starting points using the black-box approach. Here, we have opted for the Expected Improvement as the acquisition function. Since this acquisition function attempts to directly improve the current optimum every round, no adjustment is necessary after 9 optimization rounds; the final proposal is therefore also made using EI. For further details regarding its implementation, please refer to the Supporting Information.

At this point, we want to refer to Table 3 to clarify the terminology in Figure 6, where the results for the process optimization for PTQ10:BTP-eC9 are shown. Figure 6a shows the mean values of the measured efficiencies for the algorithm's proposals as a function of the optimization round. The model-based optimization approach achieves greater efficiency than the black-box approach after 10 rounds. This observation will be examined and discussed in detail below. In particular, the final round 10 will be examined in detail (Figure 6b).

For each proposal generated by the algorithm, a dozen solar cells were fabricated. The measured quantities of these solar cells must then be statistically evaluated. Defective devices are sorted out (please refer to the Supporting Information), and for the remainder, the means and variances of the respective model parameters θ_{mod} and the efficiency η are determined, which are then used for training the Gaussian processes. The case in which the variance is a function of the parameter space is called heteroscedasticity. However, this physically intuitive approach can lead to complications during the training of the surrogate model. Extremely small variances can lead to complications during model training, which then strongly affect the final surrogate model for the solar cell efficiency. For an illustrative example and a more detailed discussion, please refer to the Supporting Information.

As a result, the model-based approach appears to be significantly more prone to errors than the black-box approach, and proper model training seems to be a key factor. Various approaches are conceivable here. Adjusting the kernel function would be a plausible example. However, a particularly straightforward method in this case is to artificially inflate the variance, for instance, by calculating it homoscedastically. But also more sophisticated possibilities to inflate the variance should be explored, in particular, so as not to lose all the information contained in the measured variances. To further improve the model's accuracy, we deliberately removed outliers from the training data (see Supporting Information). In addition, we considered a combination of heteroscedastic and homoscedastic calculations. To do this, we calculated the variance σ^2 as the sum of both, and multiplied the homoscedastic component by a factor f, yielding

$$\sigma^2_{train} = \sigma^2_{het} + \sigma^2_{hom} \cdot f \tag{9}$$

The corresponding results for the process optimization using this 'mixed' approach with $f = 0.01$, $f = 0.1$, and $f = 1$ can be found in Figure 6b. It appears that higher variance, and thus smoother functions for the model parameters, yield better results. The very best result was observed for a mixed variance calculation using $f = 1$. This approach should by no means be understood as a final and optimal solution to the observed phenomenon, but rather as a simple way to improve model accuracy. Equation (9) introduces an additional hyperparameter that needs to be determined and can vary depending on the system being optimized. This is comparable to the advantages and disadvantages in UCB acquisition. While the hyperparameter increases flexibility in application, it also increases the manual complexity of the algorithm, which should be avoided. More sophisticated ways to ensure proper training of the model parameters should be part of future research.

TABLE 3 | Terminological clarification and additional explanation for Figure 6b, where the strategy for the model-based approach is to use exploitation only to find the surrogate models' final best shot. This can be understood as an Upper Confidence Bound using $\lambda = 0$. For all 'mixed' approaches, additionally, some data points were filtered out from the training data set. More information on this can be found in the main part.

Approach	Description
heteroscedastic	The assumption is that the variance of the measured data is a function of the parameter space. From a physical standpoint, in the given case, this is the intuitive approach. To illustrate this, we can consider a scenario in which certain process parameters yield a substantially higher degree of inhomogeneity in the layer thickness compared to others. In such instances, this layer thickness exhibits greater variance across the substrate. For the model training, this means a Gaussian process is trained on the data using the measured variances for each point separately. Thus, the variance is different for each point. However, for some of the model parameters (e.g., V_{OC}), extremely small variances are observed here that can be numerically problematic.
homoscedastic	It is assumed here that the variance is constant across the entire range of process parameters. For the training, this means that the variance is calculated collectively from all proposals, rather than individually for each point in the parameter space. The final variance is therefore the same for all data points. Physically speaking, this may not make much sense in this particular case, but it is a straightforward approach to inflate the measured variances to a sensible order of magnitude and avoid very small variances.
mixed-f = *	A combination of the heteroscedastic and homoscedastic approaches. The variance is calculated as a sum of homoscedastic variance and the heteroscedastic variance. The latter is multiplied by f. We tested $f = 0.01$, $f = 0.1$, and $f = 1$.

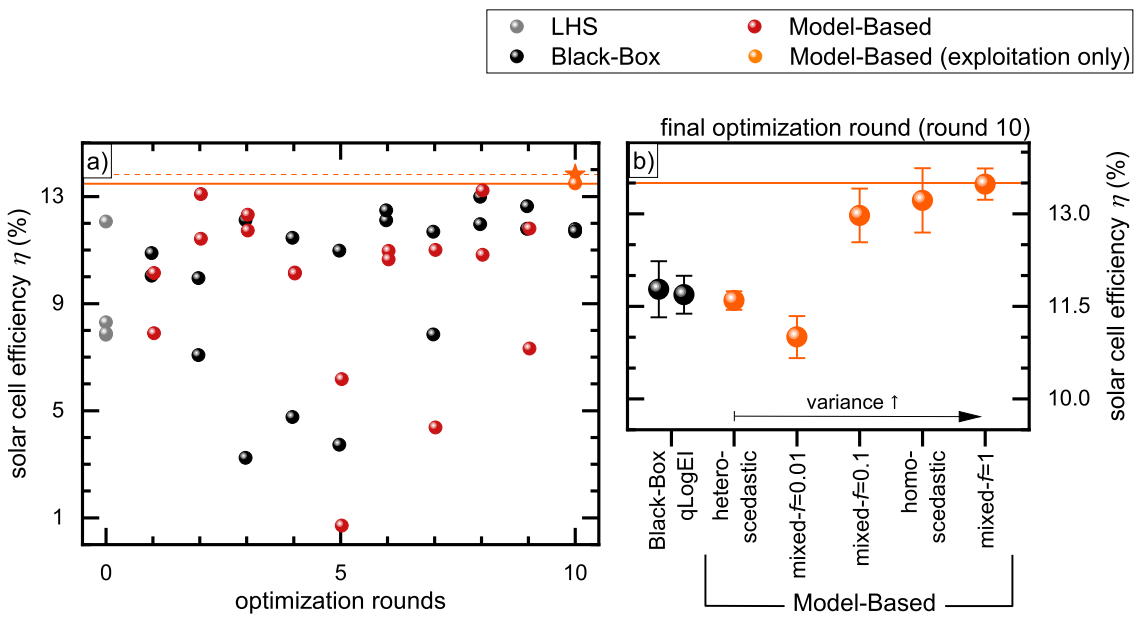

FIGURE 6 | Results for the process optimization for PTQ10:BTP-eC9 comparing the model-based approach with a black-box approach. (a) The solar cell efficiency as a function of each optimization round. Both approaches start with the same initial Latin hypercube sampling and then explore-exploit the parameter space round by round. For the model-based approach, the final optimization round (round 10) is based on exploitation only. The star and the dashed line indicate the overall best solar cell efficiency. (b) The results of the final optimization round using various surrogate models. For the black-box approach, an EI-based acquisition function from the Python package botorch, namely qLogEI, is used. For further information regarding the model-based approaches, please refer to Table 3. For the *JV* curve and the *EQE* measurements for the record cell, please refer to the Supporting Information.

4 | Statistical Validation

The results of the optimization may vary in detail with each run. In particular, the initial Latin hypercube sampling can influence the course of the optimization. Therefore, drawing statistical conclusions based on a single run is problematic. In computer science, benchmark functions are commonly used when testing optimization algorithms to solve the problem of comparing statistically significant numbers of different optimization runs that vary in their randomly generated starting points [54–56]. However, these benchmark functions are generally generic and do not by design represent the same complexity of topology of the multidimensional process parameter space as would be the case in a practical organic photovoltaics application. Furthermore,

analytical benchmark functions would make it impossible to test the model-based optimization approach. Therefore, we previously developed a benchmark function specifically intended for organic photovoltaics for precisely such cases [57]. This function is based on hundreds of experimentally characterized actual organic solar cells spanning a huge range of process parameters. The benchmark function essentially interpolates between the experimental data and creates a continuous multidimensional function that correlates process parameters with device performance and estimates of material parameters, and therefore allows testing optimization algorithms many times to generate statistically significant results on a topology of representative complexity as would be present in a real-life scenario.

Henceforth, we will refer to our benchmark function as a virtual laboratory which involves a multidimensional function approximation in which the charge carrier mobility μ, the recombination coefficient k, the active layer thickness d, which defines the generation rate $G(d)$ directly leading to $J_{gen}(d)$, analogous to as described above, and the charge-carrier dissociation efficiency η_{eh} are represented as functions of six input process parameters θ_{pro}. Based on these internal variable parameters, the *JV* curve is then simulated using drift-diffusion simulations [35, 58, 59]. Using Equation (1), the solar cell efficiency η can be directly calculated from the *JV* curve. This already provides everything needed for black-box optimization, where the selected process parameters are directly linked to the solar cell efficiency, as described above. However, for the model-based approach, the dark shunt resistance R_p is omitted here, as it is not represented within the benchmark function. As described above, this parameter is not necessary in all cases, and the benchmark function does not include cases where the dark shunt is too small. The active layer thickness d and the charge-carrier dissociation efficiency η_{eh} can be taken directly from the benchmark function. The open-circuit voltage V_{OC} is taken from the simulated *JV* curve, and the apparent ideality factor n_{app} is calculated using

$$n_{app} = 1 + \sqrt{\frac{q^2 k G d^4}{4\mu^2 (k_B T)^2}} \tag{10}$$

with the elementary charge q, the recombination constant k, the generation rate G, the active layer thickness d, the charge-carrier mobility μ, the Boltzmann constant k_B, and the temperature T [34]. Consequently, all the necessary components for the model-based approach have now been assembled.

Since the process parameter space has 6 dimensions in the case of the benchmark function, the optimization algorithm is slightly adapted compared to the experimental trials. The initial Latin hypercube sampling provides 6 points. Then, 20 optimization rounds with 4 proposals each round are performed. For a statistical analysis, 36 optimization runs are tested with different initial conditions. Furthermore, the virtual laboratory allows the topology of the parameter space to be influenced within the logic of thought experiments. For instance, if one has determined mobilities for the hundreds of solar cell *JV* curves the virtual laboratory is based on, and if that laboratory knows the device physics of solar cells, it is straightforward to vary mobilities or recombination coefficients to generate more or less optimistic versions of the virtual laboratory. For our purposes, it is useful to generate versions that have the global optimum either at the first or second interference maximum of the photocurrent vs. thickness relationship. To increase the complexity of the topology, it is useful to use those versions that peak at the second interference maximum, as this is more challenging for the algorithm to find and is more instructive for the comparison of black-box vs. model-based optimization algorithms.

Figure 7 shows the results for the different optimization approaches. We compare each optimization approach (black-box and model-based) in combination with both acquisition strategies, which we also used in the experimental validation (Expected Improvement and Knowledge Gradient). Figure 7a shows the results of the black-box approach using an EI-acquisition function, which is the most traditional approach for Bayesian Optimization. The Expected Improvement has proven to be a particularly effective acquisition function. In the case of the benchmark function used, after 20 rounds of optimization, the global optimum was found in ~1/3 of all iterations. Furthermore, it should also be noted that there was no iteration that did not result in a significant improvement compared to the start of the optimization. In Figure 7b, the black-box approach is combined with the Knowledge Gradient, and Figure 7c shows the model-based approach with EI. For a more in-depth discussion of the results, please refer to the Supporting Information. Finally, Figure 7d shows the results for the model-based approach in combination with the Knowledge Gradient. Overall, this approach yields the best results, which is also consistent with what the literature suggests [8, 49]. With a few exceptions, after 20 rounds of optimization, the global optimum is identified for ~90% of the optimization iterations.

5 | Conclusion

Efficient process optimization is urgently needed in many areas of research and industry. The use of modern optimization algorithms is key. However, valuable knowledge from recent decades of research or other potential sources of information often remains unused during process optimization. It is important to identify ways in which this knowledge can be utilized here. The approach examined in this work, a combination of Bayesian Optimization with a physical model using more sophisticated acquisition functions such as the Knowledge Gradient, appears very promising. Although there is still room for numerous improvements, we demonstrated the feasibility of the approach and directly compared it with a currently widely used optimization method. In the model-based approach, precise Gaussian processes for the individual model parameters are particularly crucial. Careful preprocessing of the data, possibly including regularization approaches such as artificially increasing the measured variances, appears to be necessary. Additional measurements of components that have not been fully processed could further increase this accuracy and improve the efficiency of the optimization.

Future research should investigate the extent to which the physical model can be supplemented or replaced to access additional sources of information, such as morphology. The annealing conditions, as well as the chemical conditions, are examples of process parameters that directly determine the morphology,

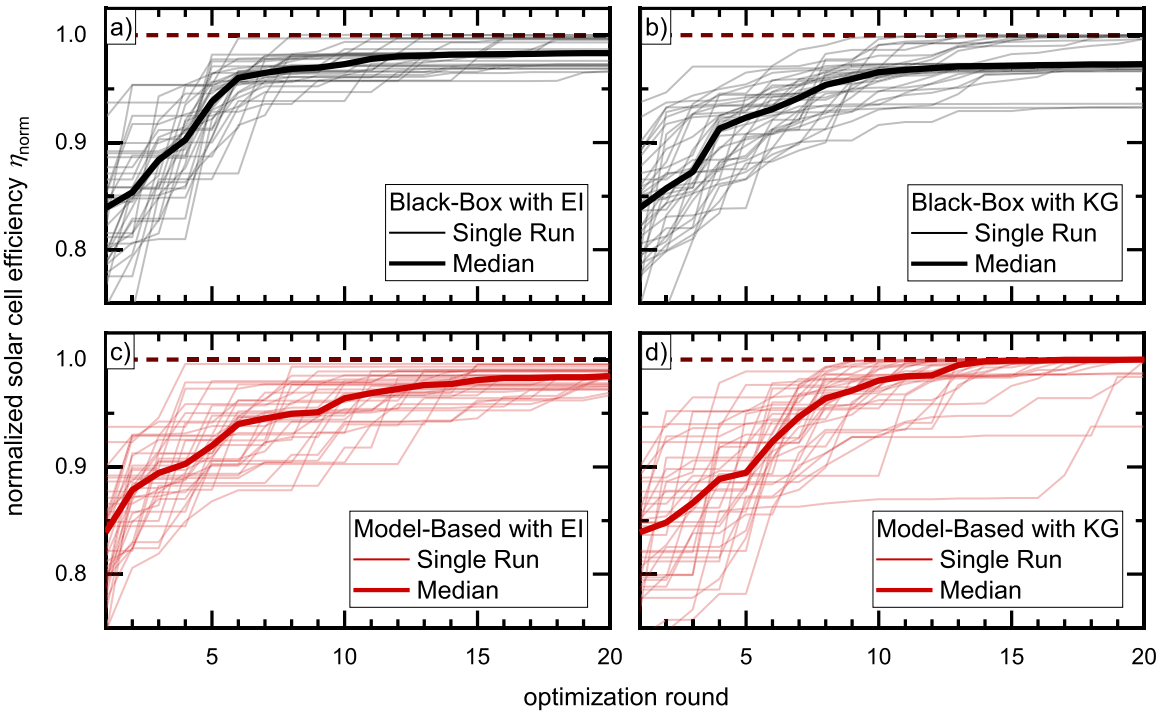

FIGURE 7 | Statistic comparison with the results from the benchmark function for (a) black-box process optimization using the EI acquisition function, (b) black-box process optimization using the KG, (c) model-based process optimization using EI, and (d) model-based process optimization using the KG strategy. The figure shows the efficiency as a function of the number of optimization rounds for a total of 36 optimization runs, as well as the median for all optimization runs. The model-based approach with the Knowledge Gradient gives the overall best results.

which directly influences charge carrier mobility, recombination effects, and thus the apparent ideality factor, or simply η_{eh} [60–64]. If these influences could also be incorporated into a physical model, process optimization could be significantly accelerated. Approaches for potential models already exist [60]. Ultimately, the key question is always whether the additional effort involved in measurement will speed up the optimization process, and whether that effort is ultimately less than what a longer optimization process would require.

Furthermore, future research should investigate which acquisition functions are particularly well-suited for the model-based approach. Less computationally intensive approaches are useful for practical applications. Thus, it should also be examined whether classical acquisition functions can be meaningfully adapted to the model-based approach; one example here is the expected improvement adapted for composite functions [27]. However, as computing power continues to grow, even more computationally intensive algorithms will eventually become practical. Information-theoretical approaches or non-myopic (multi-lookahead) approaches could be promising in this context [50, 65, 66].

Author Contributions

Leonard Christen: conceptualization, investigation, writing – original draft, methodology, software, data curation, formal analysis, visualization, validation. **Thomas Kirchartz**: supervision, funding acquisition, writing – review and editing, project administration, resources.

Acknowledgements

The authors acknowledge the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia for funding via the ENFA ("Entwicklung effizienter ternärer NFA-basierter organischer Photovoltaik durch Machine-Learning Methoden") project, the Jülich Supercomputer Centre for the granted computing time, Dr. Karsten Bittkau for conducting the ellipsometry measurements, and Sophia Lackhoff and Katharina Ortel for proofreading the manuscript and for their helpful comments.

Open access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. J. Zhang, W. Xu, S. Yang, and Y. Cui, "Development of Machine Learning in Organic Photovoltaic Cells †," *Chinese Journal of Chemistry* 44 (2026): 1239–1257, https://doi.org/10.1002/cjoc.70425.

2. H. Togun, A. Basem, M. J. Jweeg, N. Biswas, A. M. Abed, D. Paul, H. I. Mohammed, A. Chattopadhyay, B. K. Sharma, and T. Abdulrazzaq, "Advancing Organic Photovoltaic Cells for a Sustainable Future: the Role of Artificial Intelligence (AI) and Deep Learning (DL) in Enhancing Performance and Innovation," *Solar Energy* 291 (2025): 113378, https://doi.org/10.1016/j.solener.2025.113378.

3. L. A. Iturralde Carrera, G. Alfonso-Francia, C. D. Constantino-Robles, J. Terven, E. A. Chávez-Urbiola, and J. Rodríguez-Reséndiz, "Advances

and Optimization Trends in Photovoltaic Systems: a Systematic Review," *AI* 6 (2025): 225, https://doi.org/10.3390/ai6090225.

4. D. R. Ahmed and F. F. Muhammadsharif, "A Review of Machine Learning in Organic Solar Cells," *Processes* 13 (2025): 393.

5. S. Datta, A. Baul, G. C. Sarker, P. K. Sadhu, and D. R. Hodges, "A Comprehensive Review of the Application of Machine Learning in Fabrication and Implementation of Photovoltaic Systems," *IEEE Access* 11 (2023): 77750–77778, https://doi.org/10.1109/ACCESS.2023.3298542.

6. D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, and S. Wrobel, "A Review of Machine Learning for the Optimization of Production Processes," *The International Journal of Advanced Manufacturing Technology* 104 (2019): 1889–1902, https://doi.org/10.1007/s00170-019-03988-5.

7. M. Hußner, R. A. Pacalaj, G. Olaf Müller-Dieckert, C. Liu, Z. Zhou, N. Majeed, S. Greedy, I. Ramirez, N. Li, S. M. Hosseini, C. Uhrich, C. J. Brabec, J. R. Durrant, C. Deibel, and R. C. I. MacKenzie, "Machine Learning for Ultra High Throughput Screening of Organic Solar Cells: Solving the Needle in the Haystack Problem," *Advanced Energy Materials* 14 (2023): 2303000.

8. P. I. Frazier, "*A Tutorial on Bayesian Optimization, arXiv:180702811*", (2018), https://doi.org/10.48550/arXiv.1807.02811.

9. X. Liu, J. Le, C. Xie, X. Lu, R. Tian, Z. Xiong, and Y. Zhou, "Progress and Outlooks of Large-area Flexible Organic Photovoltaic Modules," *Energy & Environmental Science* 19 (2026): 791–814, https://doi.org/10.1039/D5EE06488K.

10. J. Cui, "Key Technologies and Development Trends of Building Integrated Photovoltaic (BIPV)," *International Journal of Engineering Advances* 3 (2026): 38–48, https://doi.org/10.71222/yx8vev15.

11. G. Zhang, F. R. Lin, F. Qi, T. Heumüller, A. Distler, H.-J. Egelhaaf, N. Li, P. C. Y. Chow, C. J. Brabec, A. K. Y. Jen, and H.-L. Yip, "Renewed Prospects for Organic Photovoltaics," *Chemical Reviews* 122 (2022): 14180–14274, https://doi.org/10.1021/acs.chemrev.1c00955.

12. S. Yuan, W. Luo, M. Xie, and H. Peng, "Progress in Research on Organic Photovoltaic Acceptor Materials," *RSC Advances* 15 (2025): 2470–2489.

13. Y. Zhang, H. Xia, J. Yu, Y. Yang, and G. Li, "Materials and Device Engineering Perspective: Recent Advances in Organic Photovoltaics," *Advanced Materials* 37 (2025): 2504063, https://doi.org/10.1002/adma.202504063.

14. W. Lowrie, R. J. E. Westbrook, J. Guo, H. I. Gonev, J. Marin-Beloqui, and T. M. Clarke, "Organic Photovoltaics: the Current Challenges," *The Journal of Chemical Physics* 158 (2023): 110901, https://doi.org/10.1063/5.0139457.

15. H. Lee, C. Park, D. H. Sin, J. H. Park, and K. Cho, "Recent Advances in Morphology Optimization for Organic Photovoltaics," *Advanced Materials* 30 (2018): 1800453, https://doi.org/10.1002/adma.201800453.

16. X. Yang, Y. Xie, J. Tian, J. Chen, Z. Zhang, D. Tang, X. Shi, X. Hao, J. Zhang, L. Ding, Y. Sun, and M. Lv, "Achieving a Record Fill Factor of Approaching 84% and 21% Efficiency in Binary Organic Solar Cells via Solid Additive Engineering," *Advanced Materials* 38 (2026): 19230.

17. G. Tian, Y. Chen, Y. Li, L. Liu, Q. Ma, S. Duan, C. Uragami, H. Hashimoto, P. Huang, C. Yang, Y. Yang, S. Lu, and Z. Xiao, "Ester-functionalized Nonfullerene Acceptors Modulate Crystallinity Enabling 20% Efficiency Organic Solar Cells with Scalability," *Materials Science and Engineering: R: Reports* 167 (2026): 101118, https://doi.org/10.1016/j.mser.2025.101118.

18. Y. Jiang, K. Liu, F. Liu, G. Ran, M. Wang, T. Zhang, R. Xu, H. Liu, W. Zhang, Z. Wei, Y. Cui, X. Lu, J. Hou, and X. Zhu, "20.6% Efficiency Organic Solar Cells Enabled by Incorporating a Lower Bandgap Guest Nonfullerene Acceptor without Open-Circuit Voltage Loss," *Advanced Materials* 37 (2025): 2500282, https://doi.org/10.1002/adma.202500282.

19. K. A. Mazzio and C. K. Luscombe, "The Future of Organic Photovoltaics," *Chemical Society Reviews* 44 (2015): 78–90, https://doi.org/10.1039/C4CS00227J.

20. C. M. Nkinyam, C. O. Ujah, K. C. Nnakwo, and D. V. V. Kallon, "Insight into Organic Photovoltaic Cell: Prospect and Challenges," *Unconventional Resources* 5 (2025): 100121.

21. J. Zuo, D. Han, H. Yao, V. Kuvondikov, and L. Ye, "Stretching the Future: Strategies and Emerging Trends in Stretchable Organic Photovoltaic Materials," *Energy & Environmental Science* 18 (2025): 6344–6365, https://doi.org/10.1039/D5EE01504A.

22. E. K. Solak and E. Irmak, "Advances in Organic Photovoltaic Cells: a Comprehensive Review of Materials, Technologies, and Performance," *RSC Advances* 13 (2023): 12244.

23. Y. Choi, D. Kim, S. U. Ryu, H. Kim, S. Kim, M. Kim, and T. Park, "Stability of Organic Photovoltaics: from Root Causes to Advanced Analytical Techniques," *Advanced Energy Materials* 16 (2026): 2501340, https://doi.org/10.1002/aenm.202501340.

24. B. Cao, L. A. Adutwum, A. O. Oliynyk, E. J. Luber, B. C. Olsen, A. Mar, and J. M. Buriak, "How To Optimize Materials and Devices via Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics," *ACS Nano* 12 (2018): 7434–7444, https://doi.org/10.1021/acsnano.8b04726.

25. K. Terayama, M. Sumita, R. Tamura, and K. Tsuda, "Black-Box Optimization for Automated Discovery," *Accounts of Chemical Research* 54 (2021): 1334–1346, https://doi.org/10.1021/acs.accounts.0c00713.

26. N. Asprion, R. Böttcher, R. Pack, M.-E. Stavrou, J. Höller, J. Schwientek, and M. Bortz, "Gray-Box Modeling for the Optimization of Chemical Processes," *Chemie Ingenieur Technik* 91 (2019): 305–313, https://doi.org/10.1002/cite.201800086.

27. R. Astudillo and P. Frazier, "Bayesian Optimization of Composite Functions," in International Conference on Machine Learning (PMLR, 2019), 354.

28. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed Machine Learning," *Nature Reviews Physics* 3 (2021): 422–440, https://doi.org/10.1038/s42254-021-00314-5.

29. D. Khatamsaz, R. Neuberger, A. M. Roy, S. H. Zadeh, R. Otis, and R. Arróyave, "A Physics Informed Bayesian Optimization Approach for Material Design: Application to NiTi Shape Memory Alloys," *npj Computational Materials* 9 (2023): 221, https://doi.org/10.1038/s41524-023-01173-7.

30. "Best Research-Cell Efficiency Chart National Laboratory of the Rockies", (2026): https://www.nlr.gov/pv/cell-efficiency.

31. N. Yang, S. Zhang, Y. Cui, J. Wang, S. Cheng, and J. Hou, "Molecular Design for Low-cost Organic Photovoltaic Materials," *Nature Reviews Materials* 10 (2025): 404–424, https://doi.org/10.1038/s41578-025-00792-4.

32. R. Sun, X. Yuan, X. Yang, Y. Wu, Y. Shao, X. Wu, C. J. Brabec, and J. Min, "Cost-efficient Recycling of Organic Photovoltaic Devices," *Joule* 8 (2024): 2523–2538, https://doi.org/10.1016/j.joule.2024.06.006.

33. W. Yang, W. Wang, Y. Wang, R. Sun, J. Guo, H. Li, M. Shi, J. Guo, Y. Wu, T. Wang, G. Lu, C. J. Brabec, Y. Li, and J. Min, "Balancing the Efficiency, Stability, and Cost Potential for Organic Solar Cells via a New Figure of Merit," *Joule* 5 (2021): 1209–1230, https://doi.org/10.1016/j.joule.2021.03.014.

34. D. Neher, J. Kniepert, A. Elimelech, and L. J. A. Koster, "A New Figure of Merit for Organic Solar Cells with Transport-limited Photocurrents," *Scientific Reports* 6 (2016): 24861, https://doi.org/10.1038/srep24861.

35. R. E. I. Schropp and M. Zeman, *Electrical Device Modeling* (Springer US, 1998), 117.

36. C. Wang, R. C. I. MacKenzie, U. Würfel, D. Neher, T. Kirchartz, C. Deibel, and M. Saladina, "Transport Resistance Dominates the Fill Factor Losses in Record Organic Solar Cells," *Advanced Energy Materials* 16 (2026): 2405889, https://doi.org/10.1002/aenm.202405889.

37. M. Saladina and C. Deibel, "Transport Resistance Strikes Back: Unveiling Its Impact on Fill Factor Losses in Organic Solar Cells," *Reports on Progress in Physics* 88 (2025): 038001, https://doi.org/10.1088/1361-6633/adb20c.

38. R. Garnett, *Bayesian Optimization* (Cambridge University Press, 2023), https://doi.org/10.1017/9781108348973.

39. S. Abdelghafar, H. Alshater, L. M. Abouelmagd, A. Darwish, and A. E. Hassanien, "Organic Photovoltaic Prediction Model Based on Bayesian Optimization and Explainable AI," *Scientific Reports* 15 (2025): 32559, https://doi.org/10.1038/s41598-025-18632-4.

40. K. Tsoi and S. Yerci, "Bayesian Optimization with Experience for Fast Development of Monolithic Tandem Solar Cells: Simulation Case Study," *Advanced Theory and Simulations* 7 (2024): 2301013, https://doi.org/10.1002/adts.202301013.

41. F. Kumagai, K. Gotoh, S. Miyamoto, S. Kato, K. Kutsukake, N. Usami, and Y. Kurokawa, "Bayesian Optimization of Hydrogen Plasma Treatment in Silicon Quantum Dot Multilayer and Application to Solar Cells," *Discover Nano* 18 (2023): 43, https://doi.org/10.1186/s11671-023-03821-9.

42. T. Osterrieder, F. Schmitt, L. Lüer, J. Wagner, T. Heumüller, J. Hauch, and C. J. Brabec, "Autonomous Optimization of an Organic Solar Cell in a 4-dimensional Parameter Space," *Energy & Environmental Science* 16 (2023): 3984–3993, https://doi.org/10.1039/D3EE02027D.

43. W. Xu, Z. Liu, R. T. Piper, and J. W. P. Hsu, "Bayesian Optimization of Photonic Curing Process for Flexible Perovskite Photovoltaic Devices," *Solar Energy Materials and Solar Cells* 249 (2023): 112055, https://doi.org/10.1016/j.solmat.2022.112055.

44. Q. Song, Y. Bai, and Q. Chen, "The Spring of Processing Chemistry in Perovskite Solar Cells–Bayesian Optimization," *The Journal of Physical Chemistry Letters* 13 (2022): 10741–10750, https://doi.org/10.1021/acs.jpclett.2c02635.

45. Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt, and T. Buonassisi, "Machine Learning with Knowledge Constraints for Process Optimization of Open-air Perovskite Solar Cell Manufacturing," *Joule* 6 (2022): 834–849, https://doi.org/10.1016/j.joule.2022.03.003.

46. Y. Zhang, D. W. Apley, and W. Chen, "Bayesian Optimization for Materials Design with Mixed Quantitative and Qualitative Variables," *Scientific Reports* 10 (2020): 4924, https://doi.org/10.1038/s41598-020-60652-9.

47. H. C. Herbol, W. Hu, P. Frazier, P. Clancy, and M. Poloczek, "Efficient Search of Compositional Space for Hybrid Organic–inorganic Perovskites via Bayesian Optimization," *npj Computational Materials* 4 (2018): 51, https://doi.org/10.1038/s41524-018-0106-7.

48. P. I. Frazier, W. B. Powell, and S. Dayanik, "A Knowledge-Gradient Policy for Sequential Information Collection," *SIAM Journal on Control and Optimization* 47 (2008): 2410–2439, https://doi.org/10.1137/070693424.

49. R. Astudillo and P. I. Frazier, "*Thinking inside the box: A tutorial on grey-box Bayesian optimization, in WSC '21: Proceedings of the Winter Simulation Conference*", 2 (2021): 1–15, https://doi.org/10.48550/arXiv.2201.00272.

50. P. Hennig and C. J. Schuler, "Entropy Search for Information-efficient Global Optimization," *Journal of Machine Learning Research* 13 (2012): 1809.

51. J. Ren and D. Sweet, "*Optimal Initialization of Batch Bayesian Optimization, arXiv: 240417997v1*", (2024), https://arxiv.org/abs/2404.17997.

52. M. Balandat, B. Karrer, D. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "BoTorch: a Framework for Efficient Monte-Carlo Bayesian Optimization," *Advances in Neural Information Processing Systems* 33 (2020): 21524–21538.

53. N. Zhang, K. Jiang, F. R. Lin, et al., "Morphologically Engineered Multi-component Organic Solar Cells with Stratified Donor Distribution and Alloyed Acceptors for Enhanced Efficiency and Stability," *Materials Today Energy* 42 (2024): 101548.

54. M. Jamil and X. S. Yang, "A Literature Survey of Benchmark Functions for Global Optimisation Problems," *International Journal of Mathematical Modelling and Numerical Optimisation* 4 (2013): 150, https://doi.org/10.1504/IJMMNO.2013.055204.

55. B. Y. Qu, J. J. Liang, Z. Y. Wang, Q. Chen, and P. N. Suganthan, "Novel Benchmark Functions for Continuous Multimodal Optimization with Comparative Results," *Swarm and Evolutionary Computation* 26 (2016): 23–34, https://doi.org/10.1016/j.swevo.2015.07.003.

56. K. Hussain, M. N. Mohd Salleh, S. Cheng, and R. Naseem, "Common Benchmark Functions for Metaheuristic Evaluation: a Review," *JOIV: International Journal on Informatics Visualization* 1 (2017): 218–223.

57. L. Christen, B. Urbano, A. Krause, P. Hartnagel, and T. Kirchartz, "The Value of Device Characterization for the Optimization of Organic Solar Cells," *Advanced Energy Materials* 15 (2025): 2500247, https://doi.org/10.1002/aenm.202500247.

58. B. Das, *Defect Tolerant Device Geometries for Lead-halide Perovskite Solar Cells* (Rheinisch-Westfälische Technische Hochschule Aachen, 2022).

59. P. Hartnagel, *Simulation and Characterization of Energetic Disorder and Recombination in Organic Solar Cells* (Universität Duisburg-Essen, 2023).

60. T. Wang, G. Kupgan, and J.-L. Brédas, "Organic Photovoltaics: Relating Chemical Structure, Local Morphology, and Electronic Properties," *Trends in Chemistry* 2 (2020): 535–554.

61. K. Vandewal, "Interfacial Charge Transfer States in Condensed Phase Systems," *Annual Review of Physical Chemistry* 67 (2016): 113–133, https://doi.org/10.1146/annurev-physchem-040215-112144.

62. J. Nelson, J. J. Kwiatkowski, J. Kirkpatrick, and J. M. Frost, "Modeling Charge Transport in Organic Photovoltaic Materials," *Accounts of Chemical Research* 42 (2009): 1768–1778, https://doi.org/10.1021/ar900119f.

63. S. Foster, F. Deledalle, A. Mitani, T. Kimura, K.-B. Kim, T. Okachi, T. Kirchartz, J. Oguma, K. Miyake, J. R. Durrant, S. Doi, and J. Nelson, "Electron Collection as a Limit to Polymer:PCBM Solar Cell Efficiency: Effect of Blend Microstructure on Carrier Mobility and Device Performance in PTB7:PCBM," *Advanced Energy Materials* 4 (2014): 1400311, https://doi.org/10.1002/aenm.201400311.

64. S. A. Mollinger, K. Vandewal, and A. Salleo, "Microstructural and Electronic Origins of Open-Circuit Voltage Tuning in Organic Solar Cells Based on Ternary Blends," *Advanced Energy Materials* 5 (2015): 1501335, https://doi.org/10.1002/aenm.201501335.

65. K. Qin, L. J. Hong, and W. Fan, "Non-myopic Knowledge Gradient Policy for Ranking and Selection," in *2022 Winter Simulation Conference (WSC)* (IEEE, 2022), 3051, https://doi.org/10.1109/WSC57314.2022.10015275.

66. X. Yue and R. A. L. Kontar, "Why Non-myopic Bayesian Optimization Is Promising and How Far Should We Look-ahead? A Study via Rollout," in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2020), 2808.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: aenm71249-sup-0001-SuppMat.docx.