Supporting Information

Pick a Color MARIA: Adaptive Sampling Enables the Rapid Identification of Complex Perovskite Nanocrystal Compositions with Defined Emission Characteristics

Leonard Bezinge[†], Richard M. Maceiczyk[†], Ioannis Lignos[†], Maksym V. Kovalenko[‡§*], Andrew J. deMello[†*]

† Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, 8093 Zürich, Switzerland.

‡ Institute of Inorganic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, 8093 Zürich, Switzerland.

§ Laboratory for Thin Films and Photovoltaics, Empa - Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, 8600 Dübendorf, Switzerland.

* E-mail: mvkovalenko@ethz.ch, andrew.demello@chem.ethz.ch.

S1 Regression Kriging Derivation

Maximum likelihood evaluation

The likelihood L to generate the set of observables $\mathbf{y}$ from a stochastic process with mean $\mathbf{1}\mu$ and covariance matrix $\mathrm{Cov}'(\mathbf{Y}) = \sigma^2 (\mathbf{R} + \Lambda\,\mathbf{I})$ is given by:[1]

$$L = \frac{1}{(2\pi)^{n/2}\,(\sigma^2)^{n/2}\,|\mathbf{R} + \Lambda\,\mathbf{I}|^{1/2}} \exp\left(-\frac{(\mathbf{y}-\mathbf{1}\mu)^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}(\mathbf{y}-\mathbf{1}\mu)}{2\sigma^2}\right) \tag{1}$$

We aim to find a combination of parameters (Λ, θ_1,..., θ_d) that maximizes the likelihood expression L (maximum likelihood evaluation, MLE). The log-likelihood, $\log(L)$, reads:

$$\log(L) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2}\log|(\mathbf{R}+\Lambda\,\mathbf{I})| - \frac{(\mathbf{y}-\mathbf{1}\mu)^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}(\mathbf{y}-\mathbf{1}\mu)}{2\sigma^2} \tag{2}$$

Derivatives of $\log(L)$ with respect to μ and σ^2 are equal to zero allowing us to express these constants as a function of the model parameters. We obtain the model mean ($\hat{\mu}$) from the derivative of $\log(L)$ with respect to the mean:

$$\begin{aligned}
\frac{\mathrm{d}\log(L)}{\mathrm{d}\mu} &= -\frac{1}{2\sigma^2}\cdot\frac{\mathrm{d}}{\mathrm{d}\mu}\Big((\mathbf{y}-\mathbf{1}\mu)^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}(\mathbf{y}-\mathbf{1}\mu)\Big) \\
&= -\frac{1}{2\sigma^2}\cdot\frac{\mathrm{d}}{\mathrm{d}\mu}\Big(\mathbf{y}^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}\mathbf{y} - 2\mu\cdot\mathbf{1}^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}\mathbf{y} + \mu^2\cdot\mathbf{1}^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}\mathbf{1}\Big) \\
&= -\frac{1}{2\sigma^2}\Big(-2\cdot\mathbf{1}^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}\mathbf{y} + 2\mu\cdot\mathbf{1}^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}\mathbf{1}\Big) = 0 \\
\implies \hat{\mu} &= \frac{\mathbf{1}^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}\mathbf{y}}{\mathbf{1}^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}\mathbf{1}}
\end{aligned} \tag{3}$$

In a similar fashion, we get an estimation for the model variance $\hat{\sigma}^2$:

$$\begin{aligned}
\frac{\mathrm{d}\log(L)}{\mathrm{d}\sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{(\mathbf{y}-\mathbf{1}\mu)^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}(\mathbf{y}-\mathbf{1}\mu)}{2\,(\sigma^2)^2} = 0 \\
\implies \hat{\sigma}^2 &= \frac{(\mathbf{y}-\mathbf{1}\mu)^{\mathsf{T}}(\mathbf{R}+\Lambda\,\mathbf{I})^{-1}(\mathbf{y}-\mathbf{1}\mu)}{n}
\end{aligned} \tag{4}$$

We insert these model constants into Equation 2 and ignore constant terms to compute the concentrated log-likelihood:[2]

$$\log L_c = -\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2}\log|\mathbf{R} + \boldsymbol{\Lambda}\,\mathbf{I}| \tag{5}$$

The MLE problem is thereby reduced to maximizing the concentrated log-likelihood expression $\log L_c$ with $d+1$ parameters $(\Lambda, \theta_1, ..., \theta_d)$ and returns a set of parameters that fully defines the correlation function, thus enabling the predictive capability of the model.

MLE Summary

Parameters $\Lambda, \theta_1, ..., \theta_d$ are computed to maximise the log-likelihood L_c.

$$\max_{\theta \in \mathbb{R}_+^d, \lambda \in \mathbb{R}_+} \log \mathrm{L}_c - = -\frac{n}{2} \log\left(\hat{\sigma}^2\right) - \frac{1}{2} \log |\mathbf{R} + \mathbf{\Lambda}\, \mathbf{I}|$$

$$\text{with} \qquad \hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\mu)^\mathsf{T} (\mathbf{R} + \Lambda\, \mathbf{I})^{-1} (\mathbf{y} - \mathbf{1}\mu)}{n}$$

$$\hat{\mu} = \frac{\mathbf{1}^\mathsf{T} (\mathbf{R} + \Lambda\, \mathbf{I})^{-1}\, \mathbf{y}}{\mathbf{1}^\mathsf{T} (\mathbf{R} + \Lambda\, \mathbf{I})^{-1} \mathbf{1}}$$

Best Linear Unbiased Predictor

We seek to obtain a predictor value $\hat{y}(\mathbf{x}^*)$ at an arbitrary point $\mathbf{x}^*$. To this end, we add the (unknown) predictor $\hat{y}(\mathbf{x}^*)$ to the sampled data and define an augmented data set $\widetilde{\mathbf{y}} = [\mathbf{y}^\mathsf{T},\ \hat{y}(\mathbf{x}^*)]^\mathsf{T}$.[3] The augmented correlation matrix $\widetilde{\mathbf{R}}$ is defined as:

$$\widetilde{\mathbf{R}} = \begin{pmatrix} \mathbf{R} & \mathbf{r} \\ \mathbf{r}^\mathsf{T} & 1 \end{pmatrix} \tag{6}$$

Here, the correlation vector $\mathbf{r}(\mathbf{x}^*)$ consists of elements given by $r_i(\mathbf{x}^*) = \mathrm{Corr}(Y(\mathbf{x}^*), Y(\mathbf{x}_i))$ and quantifies the correlation of an arbitrary point with the measured data points. We assume that the augmented data set is the realisation of a stochastic process with covariance matrix $\mathrm{Cov}(\widetilde{\mathbf{y}}) = \hat{\sigma}^2 \left(\widetilde{\mathbf{R}} + \Lambda\, \mathbf{I}\right)$.[3] Note that $\hat{\sigma}^2$ is a known model constant calculated from the measured data (with the MLE) and is independent of the prediction. We seek the value of $\hat{y}(\mathbf{x}^*)$ maximizing the augmented likelihood (the likelihood of generating the augmented data set). Recalling the general log-likelihood expression (Equation 2), we see that only the last term depends on $\mathbf{y}$, thus comprising the part of the augmented likelihood L_a that we aim to maximize as a function of $\hat{y}$:[3]

$$\log L_a = -\frac{1}{2\hat{\sigma}^2} (\widetilde{\mathbf{y}} - \mathbf{1}\hat{\mu})^\mathsf{T} \left(\widetilde{\mathbf{R}} + \Lambda\, \mathbf{I}\right)^{-1} (\widetilde{\mathbf{y}} - \mathbf{1}\hat{\mu}) + \text{constant terms} \tag{7}$$

$$= -\frac{1}{2\hat{\sigma}^2} \begin{pmatrix} \mathbf{y} - \mathbf{1}\hat{\mu} \\ \hat{y}(\mathbf{x}^*) - \hat{\mu} \end{pmatrix}^\mathsf{T} \begin{pmatrix} \mathbf{R} + \Lambda\, \mathbf{I} & \mathbf{r} \\ \mathbf{r}^\mathsf{T} & 1 + \Lambda \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y} - \mathbf{1}\hat{\mu} \\ \hat{y}(\mathbf{x}^*) - \hat{\mu} \end{pmatrix} \tag{8}$$

The inverse of the augmented correlation matrix can be calculated using the Banachiewicz inversion formula for an inverse of a non-singular partitioned matrix:[4]

$$\begin{pmatrix} \mathbf{R} + \Lambda \mathbf{I} & \mathbf{r} \\ \mathbf{r}^\mathsf{T} & 1 + \Lambda \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{R}_r^{-1} + \mathbf{R}_r^{-1}\mathbf{r}(1 + \Lambda - \mathbf{r}^\mathsf{T}\mathbf{R}_r^{-1}\mathbf{r})^{-1}\mathbf{r}^\mathsf{T}\mathbf{R}_r^{-1} & -\mathbf{R}_r^{-1}\mathbf{r}(1 + \Lambda - \mathbf{r}^\mathsf{T}\mathbf{R}_r^{-1}\mathbf{r})^{-1} \\ -(1 + \Lambda - \mathbf{r}^\mathsf{T}\mathbf{R}_r^{-1}\mathbf{r})^{-1}\mathbf{r}^\mathsf{T}\mathbf{R}_r^{-1} & (1 + \Lambda - \mathbf{r}^\mathsf{T}\mathbf{R}_r^{-1}\mathbf{r})^{-1} \end{pmatrix} \tag{9}$$

with $\mathbf{R}_r = \mathbf{R} + \Lambda\, \mathbf{I}$. We insert the expression of the inverted matrix in Equation 8 and proceed with multiplication. Keeping only terms that depend on $\hat{y}(\mathbf{x}^*)$, we get:

$$\log L_a = \frac{1}{2\hat{\sigma}^2 (1 + \Lambda - \mathbf{r}^\mathsf{T}\mathbf{R}_r^{-1}\mathbf{r})} \left(-(\hat{y}(\mathbf{x}^*) - \hat{\mu})^2 + (\hat{y}(\mathbf{x}^*) - \hat{\mu}) \cdot 2\, (\mathbf{y} - \mathbf{1}\hat{\mu})\, \mathbf{r}^\mathsf{T} \mathbf{R}_r^{-1}\right) \tag{10}$$

Taking the derivative of this expression with respect to $\hat{y}$ and setting it to zero, we obtain an expression for the best linear unbiaised predictor (BLUP), *i.e.* the output value with the highest likelihood:

$$\frac{\mathrm{d}\log L_a}{\mathrm{d}\hat{y}} = 0 \qquad \Longrightarrow \hat{y}\left(\mathbf{x}^*\right) = \hat{\mu} + \mathbf{r}^\mathsf{T} \, \left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1} \, \left(\mathbf{y} - \mathbf{1}\hat{\mu}\right) \tag{11}$$

In addition, we calculate the error associated to the BLUP with a more rigorous stochastic derivation,[5] accounting for sampling error, model uncertainty and uncertainty linked to the estimation of $\hat{\mu}$ from a limited set of data.

$$\hat{s}_y^2(\mathbf{x}^*) = \hat{\sigma}^2 \left(1 + \underbrace{\Lambda}_{\text{sampling}} - \underbrace{\mathbf{r}^\mathsf{T}\left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1}\mathbf{r}}_{\text{distance to data points}} + \underbrace{\frac{\left(1 - \mathbf{r}^\mathsf{T}\left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1}\mathbf{r}\right)^2}{\mathbf{1}^\mathsf{T}\left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1}\mathbf{1}}}_{\hat{\mu}\text{ estimate}} \right) \tag{12}$$

Conceptually, we explain the first two terms with the inverse of the second derivative of the augmented likelihood.[1] The absolute value of the second derivative of Equation 10 is:

$$\left|\frac{\mathsf{d}^2\log L_a}{\mathsf{d}\hat{y}^2}\right| = \frac{1}{\hat{\sigma}^2\left(1 + \Lambda - \mathbf{r}^\mathsf{T}\left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1}\mathbf{r}\right)} \tag{13}$$

If the likelihood drops quickly around the optimal predictor value, *i.e.* the absolute value of the second derivative is large, BLUP exhibits a small error. On the other hand, a flat likelihood profile around the BLUP returns a large error as variations from the predictor values are very likely.

The Kriging BLUP defines the density function f of a predicted variable Y at position $\mathbf{x}^*$. The function follows a Gaussian distribution with mean $\hat{y}\left(\mathbf{x}^*\right)$ and variance $\hat{s}_y^2(\mathbf{x}^*)$:

$$f(Y) = \frac{1}{\sqrt{2\pi\hat{s}_y^2(\mathbf{x}^*)}} \exp\left(\frac{-\left(Y(\mathbf{x}^*) - \hat{y}\left(\mathbf{x}^*\right)\right)^2}{2\,\hat{s}_y^2(\mathbf{x}^*)}\right) \tag{14}$$

Integrating f returns the probability of measuring an observable at the point $\mathbf{x}^*$ within the range of integration.

BLUP Summary

Predictors are computed at every point of the discretized parameter space. The predictor and variance at an arbitrary point $\mathbf{x}^*$ are given by:

$$\hat{y}\left(\mathbf{x}^*\right) = \hat{\mu} + \mathbf{r}^\mathsf{T} \, \left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1} \, \left(\mathbf{y} - \mathbf{1}\hat{\mu}\right)$$

$$\hat{s}_y^2(\mathbf{x}^*) = \hat{\sigma}^2\left(1 + \Lambda - \mathbf{r}^\mathsf{T}\left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1}\mathbf{r} + \frac{\left(1 - \mathbf{r}^\mathsf{T}\left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1}\mathbf{r}\right)^2}{\mathbf{1}^\mathsf{T}\left(\mathbf{R} + \Lambda\,\mathbf{I}\right)^{-1}\mathbf{1}}\right)$$

$$\text{with} \quad \mathbf{r} = \begin{pmatrix} \mathrm{Corr}\left(Y(\mathbf{x}^*)\,, Y(\mathbf{x}_1)\right) \\ \vdots \\ \mathrm{Corr}\left(Y(\mathbf{x}^*)\,, Y(\mathbf{x}_n)\right) \end{pmatrix} = \begin{pmatrix} \exp\left(-\sum_{\ell=1}^d \theta_\ell \left|x_\ell^* - x_{1\ell}\right|^{p_\ell}\right) \\ \vdots \\ \exp\left(-\sum_{\ell=1}^d \theta_\ell \left|x_\ell^* - x_{n\ell}\right|^{p_\ell}\right) \end{pmatrix}$$

Kriging-assisted sampling

In an adaptive sampling experiment, a next sample point is selected based on previous measurements. We select the point that most likely results in a targeted value y^* using Equation 8, under the hypothesis that $\hat{y}\left(\mathbf{x}^*\right) = y^*$:

$$\log L_a = -\frac{1}{2\hat{\sigma}^2} \begin{pmatrix} \mathbf{y} - \mathbf{1}\hat{\mu} \\ y^* - \hat{\mu} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \mathbf{R} + \Lambda \mathbf{I} & \mathbf{r} \\ \mathbf{r}^{\mathsf{T}} & 1 + \Lambda \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y} - \mathbf{1}\hat{\mu} \\ y^* - \hat{\mu} \end{pmatrix} \tag{15}$$

In this case, the log-likelihood is fully defined and is computed after the MLE, without necessarily evaluating the BLUP.

S2 Experimental

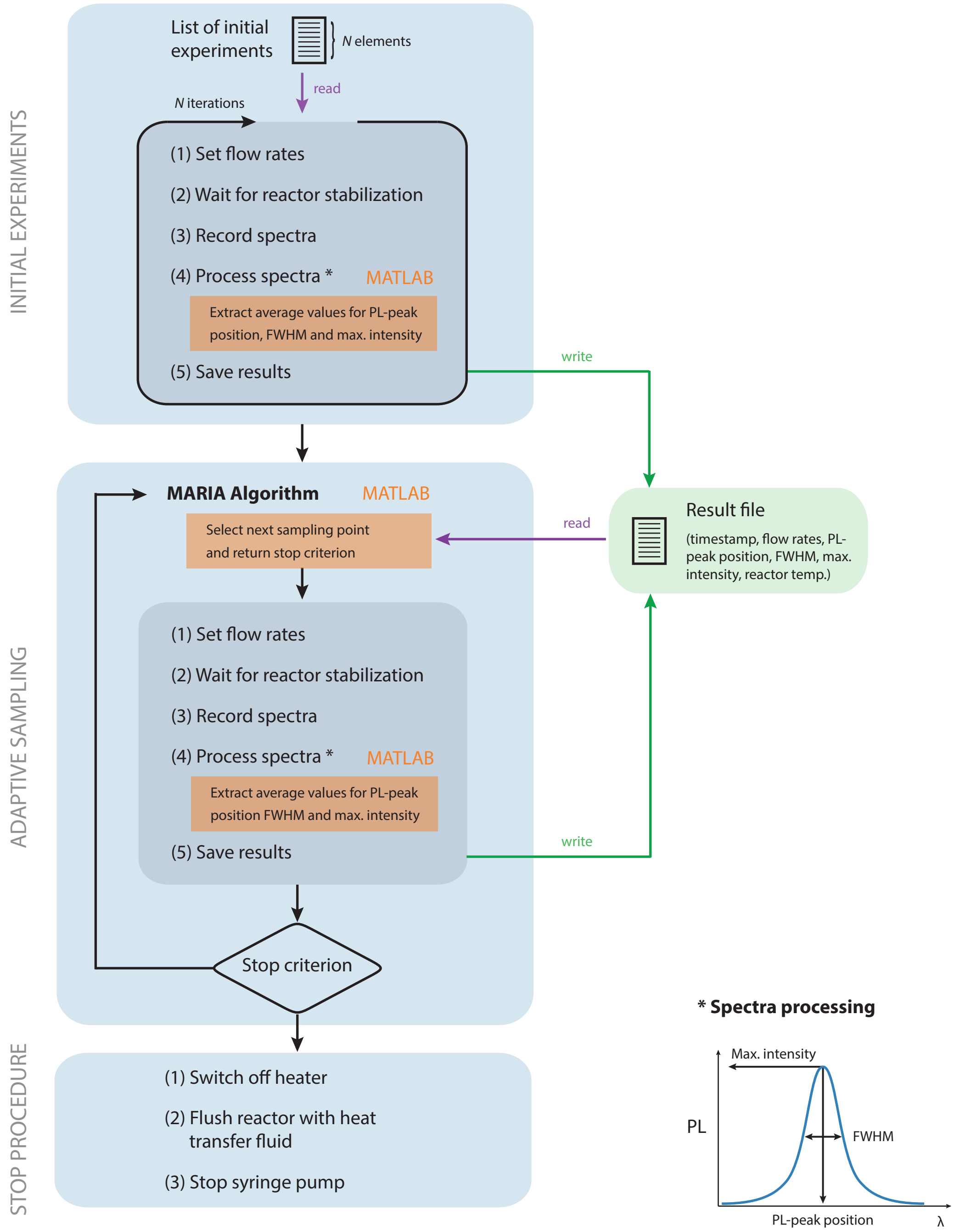

Figure S1: General structure of the LabVIEW program interfacing spectrometer, syringe pump and heater of the microfluidic reactor. Spectra processing and MARIA algorithms are implemented in MATLAB and called within the LabVIEW environment.

Reactor Stability

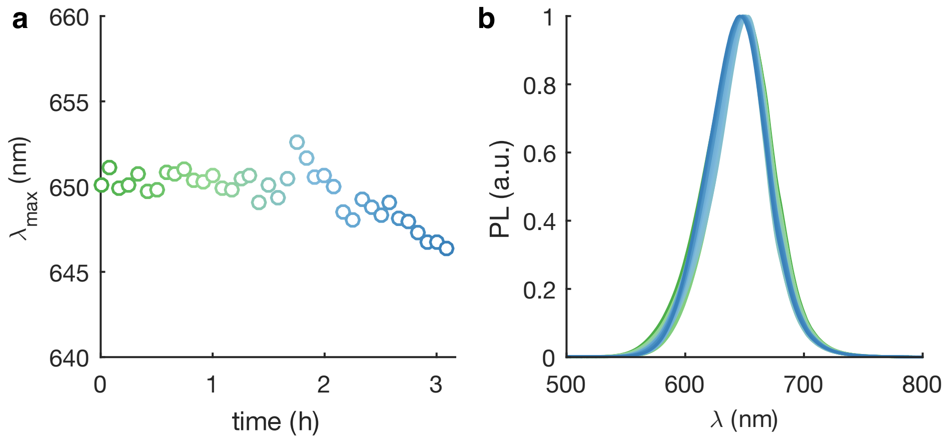

Figure S2: Reactor stability over 3 h for nanocrystals synthesized with cesium doping Cs/Pb = 0.75 and halide ratio I/(I + Br) = 0.47. PL peak wavelengths (a) are evaluated every 5 min from the measured PL spectra (b). A blue-shift of 3.7 nm is observed over the course of more than 3 h.

S3 Particle Size Distribution

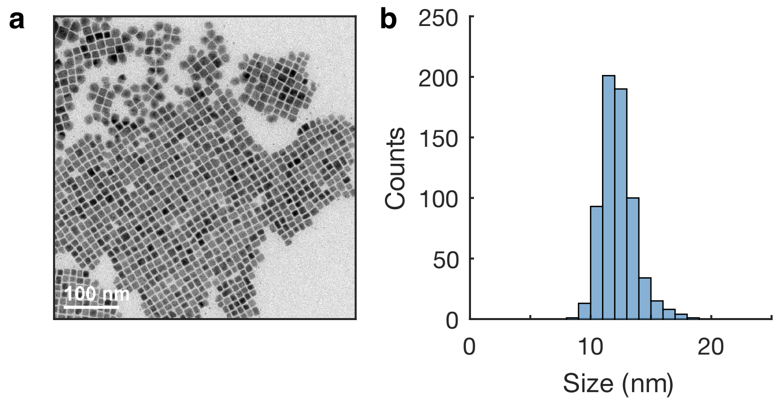

Figure S3: Representative TEM micrograph (a) of $(Cs/FA)Pb(Br/I)_3$ nanocrystals synthesised in the microfluidic reactor with the corresponding size distribution (b) determined by image analysis.

S4 Adaptive Sampling Compared to Systematic Screening

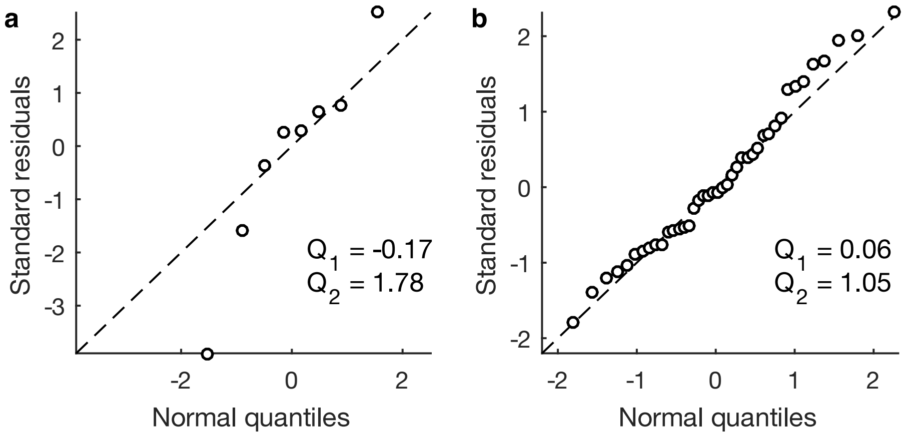

Figure S4: q-q plots of residuals obtained by leave-one-out cross-validation. The validation procedure is applied to data points with PL peak in the range 660 to 700 nm. Kriging models after 56 measurements are compared for a systematic screening approach (a) or using the MARIA procedure with a 680 nm target value (b). Statistical indicators Q_1 and Q_2 are computed in the two cases and tend to 0 and 1, respectively, for normally distributed residuals. With a larger number of measurement in the proximity of the target output, adaptive sampling produces a more robust response surface model.

S5 Response Surface Sensitivity

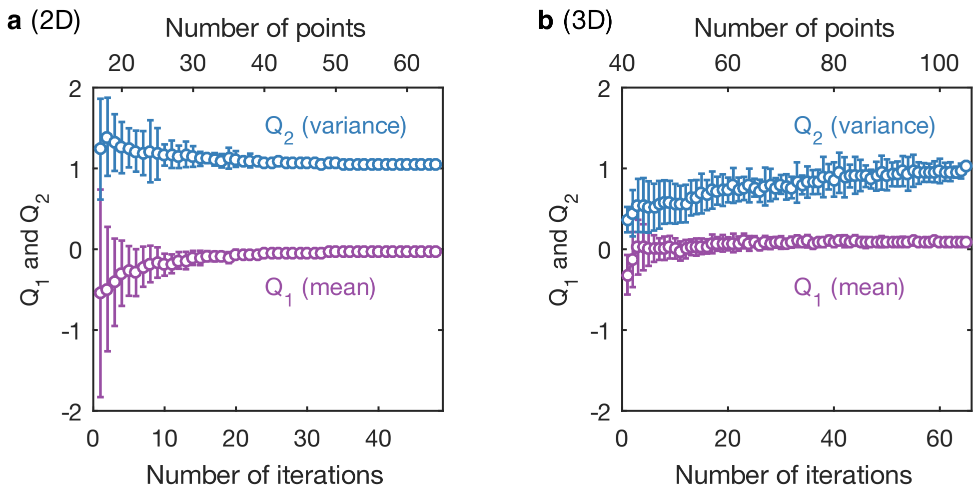

Figure S5: First and second moments (Q_1 and Q_2) of standardized residuals obtained by leave-one-out cross-validation for randomly-selected sets of measurements with increasing lengths. A robust model with normally distributed residuals is characterized by values of $Q_1 = 0$ and $Q_2 = 1$. Adaptive sampling procedures are performed in a two-dimensional parameter space for the synthesis of $(Cs/FA)Pb(Br/I)_3$ nanocrystals as shown in (a) and in three-dimensional parameter space for the synthesis of $(Rb/Cs/FA)Pb(Br/I)_3$ nanocrystals as shown in (b).

References

(1) Jones, D. R.; Schonlau, M.; Welch, W. J. Efficient Global Optimization of Expensive Black-Box Functions. *J. Glob. Optim.* **1998**, *13*, 455–492.

(2) Jones, D. R. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *J. Glob. Optim.* **2001**, *21*, 345–383.

(3) Forrester, A. I. J.; Keane, A. J.; Bressloff, N. W. Design and Analysis of "Noisy" Computer Experiments. *AIAA J.* **2006**, *44*, 2331–2339.

(4) Tian, Y.; Takane, Y. The Inverse of Any Two-By-Two Nonsingular Partitioned Matrix and Three Matrix Inverse Completion Problems. *Comput. Math. with Appl.* **2009**, *57*, 1294–1304.

(5) Sacks, J.; Welch, W. J.; Mitchell, T. J.; Wynn, H. P. Design and Analysis of Computer Experiments. *Stat. Sci.* **1989**, *4*, 409–423.