Discovering the Underlying Equations Governing Perovskite Solar-Cell Degradation Using Scientific Machine Learning

Richa R. Naik [*†] **Armi Tiihonen** [†] **Janak Thapa** [†] **Clio Batali** [‡] **Shijing Sun**[†] **Zhe Liu**[†]

Tonio Buonassisi[†]
[†]Department of Mechanical Engineering,
[‡]Department of Materials Science and Engineering
Massachusetts Institute of Technology
77 Massachusetts Ave, Cambridge, MA 02139, USA
naikr@mit.edu

Abstract

To date, most applications of machine learning (ML) in materials science have focused on process optimization and improving engineering parameters of merit. Learning physical laws directly from data has a great potential to accelerate materials research by allowing us to extract generalizable scientific information directly from data. In this study, we apply scientific ML — a blend of traditional scientific mechanistic modeling (differential equations) with machine learning methodologies — to identify equations governing the degradation of methylammonium lead iodide perovskite (MAPI). The environmental instability of halide perovskite materials is a major issue hindering the commercialization of perovskite solar cells, that have potential to provide high-performing and cost-effective solar energy in future. Discovering the underlying equations directly from perovskite degradation data could accelerate the development of stable perovskite photovoltaic technology. We aim to study the quantitative effect of temperature on MAPI decomposition with degradation at a particular temperature as the first step. We obtain high temporal resolution data describing MAPI film degradation and synthesize noise-free simulated data based on it to analyze how well the underlying differential equation can be recognized using scientific ML. This is done by applying sparse regression method PDE-FIND [17] on the simulated data. In order to investigate the robustness of the identification of governing differential equations with respect to noise, we apply varying levels of artificial noise to the simulated data. We obtain about 6% error between the exact solution and the solution identified by PDE-FIND, even when 4% Gaussian noise is added. This study demonstrates the application of scientific ML in practical materials science systems, highlighting the promise and challenges associated with ML-aided scientific discovery.

1 Introduction

The decomposition of MAPI is a complex phenomenon affected by several factors such as film properties, temperature [9, 5, 6], humidity [22, 10, 8] and illumination [14, 11]. MAPI has multiple possible reaction pathways and degrades to PbI_2 via reaction [1]:

[*]Presenting Author

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

$$\mathrm{MAPbI_3} \rightarrow \mathrm{PbI_2} + [\mathrm{CH_3NH_3}^+ + \mathrm{I}^-] \rightarrow \mathrm{PbI_2} + \mathrm{CH_3\,NH_2} + \mathrm{HI}$$

Smecca et al. [18] prove that the rate of MAPI degradation obeys an Arrhenius-type law. Using transition state theory, Fan et al. [7] approximate the surface degradation rate and find that it matches the rate observed in experiments. These studies suggest the dependence of the rate of MAPI degradation on an Arrhenius-type equation. The derivation of this dependence through first principles is difficult because of the complexity of MAPI decomposition, despite the availability of well-resolved dynamical data.

Scientific ML methods are well-suited to identify governing equations from data, especially when the systems being studied are too complicated to yield to traditional theoretical analysis. What sets scientific ML apart from traditional curve-fitting or regression is its generalizability. In order to perform curve-fitting, information on the equation form is required. Scientific ML can be used to uncover the governing equation without the knowledge of the equation form. The knowledge of underlying laws can be used to extrapolate beyond the dataset more efficiently. It can also provide key information about the physical phenomena. There are multiple methods in literature: Rudy et al. [17] describe PDE-FIND, a sparse regression approach for discovery of physical laws describing dynamical systems. First, a library of potential candidate functions is built. Differentials are calculated by finite difference or polynomial interpolation. Once a large matrix with all candidate functions is composed, different sparse regression methods may be used to extract the partial differential equation describing the system. The sparse methods implemented are sequential threshold ridge regression, lasso regression, elastic net regression and greedy algorithm. Another sparse technique is Sparse Identification of nonlinear Dynamics (SINDy) [4]. It uses a custom deep autoencoder to find a coordinate system in which the dynamics of the system are sparse, and then uses sparse regression to find the governing equations in the associated coordinate system. Atkinson et al. [2] present a generalized method for the discovery of differential equations using genetic programming. Physics Informed Neural Networks (PINN) [16] and PDE-NET [12, 13] are deep learning methodologies to extract governing partial differential equations using dynamical data.

In this study, we focus on the application of PDE-FIND to perovskite degradation data. Identifying governing differential equations directly from the experimental aging test data would provide tools for reliable lifetime prediction of perovskite solar cells. A deeper understanding of thermal degradation would help in the determination of acceleration factors for long-term aging tests. These developments would spur the advancement of the perovskite photovoltaic technology and have been called for by the community.

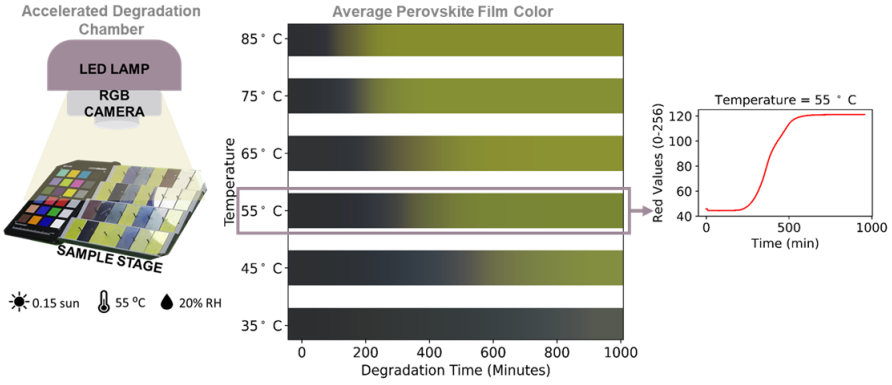


Figure 1: The experimental process. a) A schematic diagram of the in-house accelerated degradation chamber with a superimposed camera image of degrading MAPI films. b) Average perovskite film color as a function of time at different temperatures. c) Processed average red color component as a function of time for films degraded at T = 55 °C.

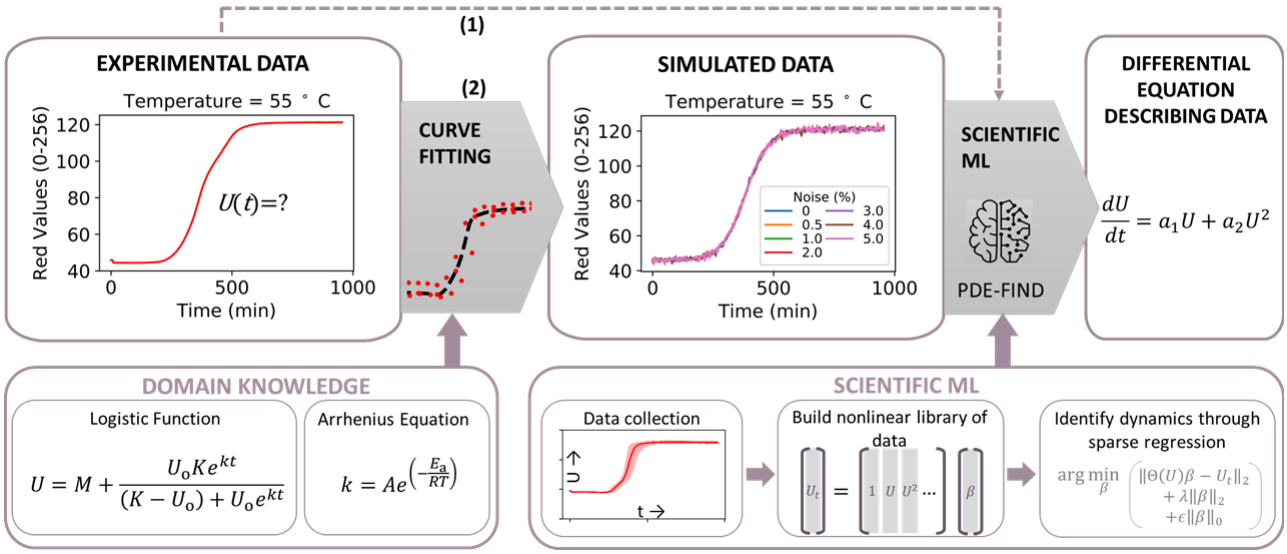


Figure 2: Schematic diagram of the data-management workflow used in this study. Workflow (1) involves the application of PDE-FIND to experimental data; Workflow (2) first fits the experimental data to create a simulated dataset, optionally adds Gaussian noise, then applies PDE-FIND

2 Methods

Our experimental workflow is shown in **Figure 1**. 140 thin-film samples of methylammonium lead iodide (MAPI) were subjected to degradation at 0.15 Sun illumination, 20% relative humidity, and temperatures varying from 35 to 85 °C in our in-house environmental chamber (Figure 1a, the in-house environmental chamber is described in greater detail in [19]). We monitored the degradation of MAPI based on the color change of the material. As MAPI films decompose, they change their color from black (100% MAPI) to yellow (0% MAPI). We acquired images of the degrading films with high temporal resolution and processed them to obtain the average red, blue and green color components of films as a function of time (Figure 1b). We use the red color time-series in the study because it captures the temporal perovskite decomposition behavior, as shown in [20](Figure 1c).

We use the non-linear least-squares method to fit our experimental data to the Verhulst logistic equation [21] to model the S-shaped curve of the change in red color as a function of time. This is a reasonable assumption because in the reactions involving the nucleation and growth of a new phase, the fraction of the new phase follows a sigmoidal curve over time[3].

$$U(t,T) = M + \frac{U_o K e^{kt}}{(K - U_o) + U_o e^{kt}} \tag{1}$$

$$\frac{\partial U(t,T)}{\partial t} = k(T)\, U(t,T) \left(1 - \frac{U(t,T)}{K}\right) \tag{2}$$

Where U_o is the initial concentration, k is growth rate and K is the carrying capacity. In the context of MAPI degradation, U_o and K can be considered arbitrary parameters. We assume k varies with temperature according to the Arrhenius equation[18]:

$$k(T) = Ae^{\left(-\frac{E_a}{RT}\right)} \tag{3}$$

We use this model to produce noise-free simulated data as well as simulated data with Gaussian noise. Our data-processing workflow is illustrated in **Figure 2**.

We apply the sparse regression methodology PDE-FIND [17] to experimental data (Workflow **(1)**). First, we use the time-series from all the temperatures to discover the partial differential equation

(PDE) defining the relationship between MAPI degradation, temperature and time. Then, we apply PDE-FIND to the degradation data at each temperature, to discover the ordinary differential equation (ODE) that describes MAPI decomposition at a particular temperature. We also apply PDE-FIND to simulated data with and without noise (Workflow **(2)**). The library of potential candidate functions consists of polynomials and other non-linear functions of U. Differentials are calculated by finite difference with convolutional smoothing using a 1D Gaussian kernel. We also apply the Savitzky–Golay filter to the time derivative to prevent the propagation and amplification of noise. Once a large matrix with all candidate functions is composed, we use sequential threshold ridge regression to identify which terms contribute to the dynamics described by the data as well as those terms' weights. The goal of this method is to find a sparse coefficient vector β that only consists of the active features that best represent the dynamics. The rest of the features are hard-thresholded to zero. The loss function we use is as follows (λ_2 and λ_0 are the L2 and L0 regularization penalties respectively):

$$\hat{\beta} = \arg\min_{\beta} \|\Theta(U)\beta - U_t\|_2 + \lambda_2\|\beta\|_2 + \lambda_0\|\beta\|_0 \tag{4}$$

3 Results

Our aim is to obtain the equation that describes the environmental degradation of methylammonium lead iodide (MAPI) as a function of time and temperature. There are two main challenges for scientific ML in this application that are common also with many other experimental applications: The function space that could in principle capture the degradation processes is enormous, complicating identification of the equations. Furthermore, experimental data is noisy, making the identification of quantitative analytic descriptions even more challenging.

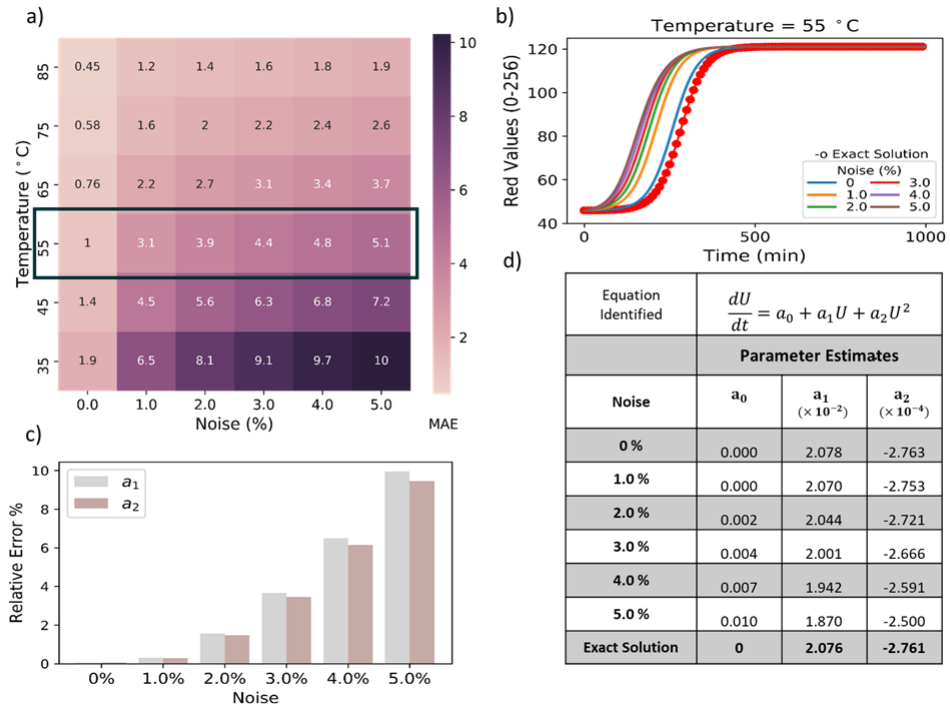
We attempt to uncover the differential equation governing perovskite degradation directly from experimental data (Workflow (1). We apply sparse regression with a broad, more general function library consisting of polynomials of U and other non-linear functions such as $\sin(U)$, $\cos(U)$ and $\exp(U)$. First, we apply PDE-FIND to the data at all temperatures to discover the governing PDE. Second, we apply PDE-FIND to the time-series at each temperature separately to identify the governing ODE. In both cases, the equation identified by PDE-FIND does not to fit the data. We apply domain knowledge to narrow the candidate function library to polynomials of U. With these constraints, we find that PDE-FIND is able to identify an ODE that fits the experimental data at each temperature. There is significant variation in the value of the coefficients accompanying the selected function terms with different experimental data points, however.

Instead of applying scientific machine learning to the experimental data directly, we combine intuition along with our knowledge of the physical system to develop a model to produce noise-free simulated data and simulated data with Gaussian noise (Workflow (2)). We use the non-linear least-squares method to fit our experimental data to the Verhulst logistic equation [21] to model the S-shaped curve of the change in red color as a function of time. We assume k varies with temperature according to the Arrhenius equation.

With the noise-free data that we simulated using the Verhulst logistic equation model, PDE-FIND estimates the underlying differential equation with significant accuracy, as shown in **Figure 3**. At 0% noise, the MAE between the exact curve and one obtained from integrating the differential equation identified by PDE-FIND is less than 2. In order to understand the performance of the method in the presence of noise, we add artificial Gaussian noise to the simulated data. At 5% Gaussian noise, PDE-FIND correctly identifies the functional terms, but the error in the fitting parameters increases to almost 10%, and the resulting integrated curve has a 5 MAE relative to the underlying noise-free simulated curve (Figure 3a). The equation identified by PDE-FIND has the same S-shaped form as the exact data (Figure 3b). We compare the error in estimating the parameter values in the differential equation describing the simulated data. We find that at 5% Gaussian noise, the relative error in parameter estimates is less than 10% for 55 °C data (Figure 3c, Figure 3d).

We also compare the functioning of equation extraction via PDE-FIND on simulated datasets at different temperatures. The Verhulst logistic equation model becomes increasingly steep and shifts to the left with higher temperature. PDE-FIND successfully identifies this trend. It appears that the MAE is higher for equation extraction at lower temperature data. On investigating further, we are able to determine that most of the error comes from fitting the rising portion of the curve. When curves at

different temperatures are appropriately truncated, the MAE at all temperatures and a particular noise level is the same.



Equation Identified	$\frac{dU}{dt} = a_0 + a_1 U + a_2 U^2$		
	Parameter Estimates		
Noise	$\mathbf{a_0}$	$\mathbf{a_1}$ ($\times 10^{-2}$)	$\mathbf{a_2}$ ($\times 10^{-4}$)
0 %	0.000	2.078	-2.763
1.0 %	0.000	2.070	-2.753
2.0 %	0.002	2.044	-2.721
3.0 %	0.004	2.001	-2.666
4.0 %	0.007	1.942	-2.591
5.0 %	0.010	1.870	-2.500
Exact Solution	**0**	**2.076**	**-2.761**

Figure 3: PDE-FIND Results: a) Heatmap of the mean absolute error (MAE) between the exact solution and the solution obtained by integrating the DE identified by PDE-FIND at different noise levels at different temperatures. b) Comparison of the exact solution at T=55 °C with solution curves obtained by integrating the DE identified by PDE-FIND. c) Relative error between the exact parameter weight and those estimated by PDE-FIND at different noise levels for T=55 °C data. d) The equation identified by PDE-FIND along with parameter weights in a tabulated format for T=55 °C and different noise levels. The last row of the table has the values of parameters for the exact solution.

4 Conclusion

We conclude that the study of materials degradation can be rendered more quantitative through the identification of differential equations that govern the process. Our approach has the potential to accelerate the understanding of materials degradation and the reliability optimization of perovskite materials. Extracting physical laws would facilitate the definition of acceleration factors for aging tests and also help in the prediction of perovskite solar cell degradation under varying environmental conditions. We demonstrate the application of a scientific ML tool, PDE-FIND on MAPI degradation data. Our results show that PDE-FIND succeeds at identifying the differential equation describing the simulated data when up to 5% Gaussian noise is added. The error in the parameter values is 10% at 5% Gaussian noise. Scientific ML methods can be immensely useful at uncovering the governing equations of dynamical systems, if the data obtained has low noise or can be denoised by noise-reduction techniques. However, most experimental data is typically noisy, and denoising the data adequately can be challenging. Our contribution motivates the development of scientific ML techniques that are more robust to noise.

Broader Impact

There remain many complex systems that have eluded quantitative analytic descriptions or even characterization of a suitable choice of variables in many disciplines such as biology, finance and materials science. With today's state-of-the art equipment, acquiring large quantities of data has never been easier. As put by Rackauckas et al. [15], *the well-known adage "a picture is worth a thousand words" might well be "a model is worth a thousand datasets."*. Scientific ML is a promising method that can be used to uncover governing equations through data, especially when the derivation of physical laws using first principles is challenging. Not only does scientific machine learning aide us with understanding the underlying scientific phenomena better, it also helps to make simulations faster and extrapolate beyond our dataset. Through our study we show that, scientific ML, in its current state, is well-suited to be applied to domains where obtaining large quantities of low-noise data is possible, and will find more applications with methods that are robust to noise.

Acknowledgments and Disclosure of Funding

The authors thank Kathleen Champion, Samuel Rudy, Zichao Long and Steven Atkinson for helpful discussions regarding scientific ML. This work was supported by DARPA (R.N.), TOTAL (A.T., S.S., Z.L.), and the U.S. Department of Energy (DOE) under Photovoltaic Research and Development (PVRD) program under Award no. DE-EE0007535 (Z.L.). S.S. and Z.L. acknowledge partial support from TOTAL Energy Fellowship through the MIT Energy Initiative. A.T. acknowledges the Alfred Kordelin Foundation.

References

[1] G. Abdelmageed, C. Mackeen, K. Hellier, L. Jewell, L. Seymour, M. Tingwald, F. Bridges, J. Z. Zhang, and S. Carter. Effect of temperature on light induced degradation in methylammonium lead iodide perovskite thin films and solar cells. *Solar Energy Materials and Solar Cells*, 174: 566–571, 1 2018. ISSN 09270248. doi: 10.1016/j.solmat.2017.09.053.

[2] S. Atkinson, W. Subber, L. Wang, G. Khan, P. Hawi, and R. Ghanem. Data-driven discovery of free-form governing differential equations. 9 2019. URL `http://arxiv.org/abs/1910.05117`.

[3] J. P. Bastos, G. Uytterhoeven, W. Qiu, U. W. Paetzold, D. Cheyns, S. Surana, J. Rivas, M. Jaysankar, W. Song, T. Aernouts, J. Poortmans, and R. Gehlhaar. Model for the prediction of the lifetime and energy yield of methyl ammonium lead iodide perovskite solar cells at elevated temperatures. *ACS Applied Materials and Interfaces*, 11:16517–16526, 5 2019. ISSN 19448252. doi: 10.1021/acsami.9b00923. URL `www.acsami.org`.

[4] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences of the United States of America*, 116:22445–22451, 11 2019. ISSN 10916490. doi: 10.1073/pnas.1906995116. URL `https://arxiv.org/abs/1904.02107v2`.

[5] B. Conings, J. Drijkoningen, N. Gauquelin, A. Babayigit, J. D'Haen, L. D'Olieslaeger, A. Ethirajan, J. Verbeeck, J. Manca, E. Mosconi, F. D. Angelis, and H. G. Boyen. Intrinsic thermal instability of methylammonium lead trihalide perovskite. *Advanced Energy Materials*, 5, 8 2015. ISSN 16146840. doi: 10.1002/aenm.201500477. URL `https://onlinelibrary.wiley.com/doi/full/10.1002/aenm.201500477https://onlinelibrary.wiley.com/doi/abs/10.1002/aenm.201500477https://onlinelibrary.wiley.com/doi/10.1002/aenm.201500477`.

[6] G. Divitini, S. Cacovich, F. Matteocci, L. Cinà, A. D. Carlo, and C. Ducati. In situ observation of heat-induced degradation of perovskite solar cells. *Nature Energy*, 1:1–6, 1 2016. ISSN 20587546. doi: 10.1038/NENERGY.2015.12. URL `www.nature.com/natureenergy`.

[7] Z. Fan, H. Xiao, Y. Wang, Z. Zhao, Z. Lin, H. C. Cheng, S. J. Lee, G. Wang, Z. Feng, W. A. Goddard, Y. Huang, and X. Duan. Layer-by-layer degradation of methylammonium lead tri-iodide perovskite microplates. *Joule*, 1:548–562, 11 2017. ISSN 25424351. doi: 10.1016/j.joule.2017.08.005.

[8] Y. Han, S. Meyer, Y. Dkhissi, K. Weber, J. M. Pringle, U. Bach, L. Spiccia, and Y. B. Cheng. Degradation observations of encapsulated planar ch3nh3pbi3 perovskite solar cells at high temperatures and humidity. *Journal of Materials Chemistry A*, 3:8139–8147, 4 2015. ISSN 20507496. doi: 10.1039/c5ta00358j. URL `www.rsc.org/MaterialsA`.

[9] E. J. Juarez-Perez, L. K. Ono, M. Maeda, Y. Jiang, Z. Hawash, and Y. Qi. Photodecomposition and thermal decomposition in methylammonium halide lead perovskites and inferred design principles to increase photovoltaic device stability. *Journal of Materials Chemistry A*, 6:9604–9612, 5 2018. ISSN 20507496. doi: 10.1039/c8ta03501f. URL `https://pubs.rsc.org/en/content/articlehtml/2018/ta/c8ta03501fhttps://pubs.rsc.org/en/content/articlelanding/2018/ta/c8ta03501f`.

[10] N. K. Kim, Y. H. Min, S. Noh, E. Cho, G. Jeong, M. Joo, S. W. Ahn, J. S. Lee, S. Kim, K. Ihm, H. Ahn, Y. Kang, H. S. Lee, and D. Kim. Investigation of thermally induced degradation in ch3nh3pbi3 perovskite solar cells using in-situ synchrotron radiation analysis. *Scientific Reports*, 7:1–9, 12 2017. ISSN 20452322. doi: 10.1038/s41598-017-04690-w. URL `www.nature.com/scientificreports`.

[11] S. W. Lee, S. Kim, S. Bae, K. Cho, T. Chung, L. E. Mundt, S. Lee, S. Park, H. Park, M. C. Schubert, S. W. Glunz, Y. Ko, Y. Jun, Y. Kang, H. S. Lee, and D. Kim. Uv degradation and recovery of perovskite solar cells. *Scientific Reports*, 6:1–10, 12 2016. ISSN 20452322. doi: 10.1038/srep38150. URL `www.nature.com/scientificreports`.

[12] Z. Long, Y. Lu, X. Ma, and B. Dong. Pde-net: Learning pdes from data. *35th International Conference on Machine Learning, ICML 2018*, 7:5067–5078, 10 2017. URL http://arxiv.org/abs/1710.09668.

[13] Z. Long, Y. Lu, and B. Dong. Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399, 11 2018. doi: 10.1016/j.jcp.2019.108925. URL http://arxiv.org/abs/1812.04426http://dx.doi.org/10.1016/j.jcp.2019.108925.

[14] W. Nie, J. C. Blancon, A. J. Neukirch, K. Appavoo, H. Tsai, M. Chhowalla, M. A. Alam, M. Y. Sfeir, C. Katan, J. Even, S. Tretiak, J. J. Crochet, G. Gupta, and A. D. Mohite. Light-activated photocurrent degradation and self-healing in perovskite solar cells. *Nature Communications*, 7:1–9, 5 2016. ISSN 20411723. doi: 10.1038/ncomms11574. URL www.nature.com/naturecommunications.

[15] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman. Universal differential equations for scientific machine learning. 1 2020. URL http://arxiv.org/abs/2001.04385.

[16] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2 2019. ISSN 10902716. doi: 10.1016/j.jcp.2018.10.045.

[17] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3:e1602614, 4 2017. ISSN 23752548. doi: 10.1126/sciadv.1602614. URL http://advances.sciencemag.org/.

[18] E. Smecca, Y. Numata, I. Deretzis, G. Pellegrino, S. Boninelli, T. Miyasaka, A. L. Magna, and A. Alberti. Stability of solution-processed mapbi3 and fapbi3 layers. *Physical Chemistry Chemical Physics*, 18:13413–13422, 5 2016. ISSN 14639076. doi: 10.1039/c6cp00721j. URL www.rsc.org/pccp.

[19] S. Sun, A. Tiihonen, F. Oviedo, Z. Liu, J. Thapa, N. Titan, P. Hartono, A. Goyal, C. Batali, A. Encinas, J. Yoo, R. Li, Z. Ren, M. Bawendi, V. Stevanovic, J. Fisher, and T. Buonassisi. A physical data fusion approach to optimize compositional stability of halide perovskites. 2020. doi: 10.26434/chemrxiv.12601997.v1. URL https://doi.org/10.26434/chemrxiv.12601997.v1.

[20] N. Titan, P. Hartono, J. Thapa, A. Tiihonen, F. Oviedo, C. Batali, J. Yoo, Z. Liu, R. Li, D. F. Marrón, M. Bawendi, T. Buonassisi, and S. Sun. How machine learning can help select capping layers to suppress perovskite degradation. *chemrxiv.org*. doi: 10.26434/chemrxiv.11766528.v1. URL https://doi.org/10.26434/chemrxiv.11766528.v1.

[21] A. Tsoularis and J. Wallace. Analysis of logistic growth models. *Mathematical Biosciences*, 179:21–55, 7 2002. ISSN 00255564. doi: 10.1016/S0025-5564(02)00096-2.

[22] J. Yang, B. D. Siempelkamp, D. Liu, and T. L. Kelly. Investigation of ch3nh3pbi3degradation rates and mechanisms in controlled humidity environments using in situ techniques. *ACS Nano*, 9:1955–1963, 2 2015. ISSN 1936086X. doi: 10.1021/nn506864k. URL www.acsnano.org1955.