







# Stability follows efficiency based on the analysis of a large perovskite solar cells ageing dataset

Received: 2 February 2023

Accepted: 1 August 2023

Published online: 12 August 2023

 Check for updates

Noor Titan Putri Hartono <sup>1,3</sup> ✉, Hans Köbler <sup>1,3</sup>, Paolo Graniero <sup>1,2</sup>, Mark Khenkin <sup>1</sup>, Rutger Schlatmann <sup>1</sup>, Carolin Ulbrich <sup>1</sup> & Antonio Abate <sup>1</sup> ✉

While perovskite solar cells have reached competitive efficiency values during the last decade, stability issues remain a critical challenge to be addressed for pushing this technology towards commercialisation. In this study, we analyse a large homogeneous dataset of Maximum Power Point Tracking (MPPT) operational ageing data that we collected with a custom-built High-throughput Ageing System in the past 3 years. In total, 2,245 MPPT ageing curves are analysed which were obtained under controlled conditions (continuous illumination, controlled temperature and atmosphere) from devices comprising various lead-halide perovskite absorbers, charge selective layers, contact layers, and architectures. In a high-level statistical analysis, we find a correlation between the maximum reached power conversion efficiency (PCE) and the relative PCE loss observed after 150-hours of ageing, with more efficient cells statistically also showing higher stability. Additionally, using the unsupervised machine learning method self-organising map, we cluster this dataset based on the degradation curve shapes. We find a correlation between the frequency of particular shapes of degradation curves and the maximum reached PCE.

Perovskite solar cells (PSCs) have reached a competitive efficiency of 26.1%<sup>1</sup>, indicating that the technology has the potential to be commercialised and implemented on a large scale. However, the current PSCs lifetime is subpar (-1 order of magnitude lower) compared to silicon solar cells<sup>2</sup>, even if environmental stressors like water and oxygen are excluded.

Several publications have intended to gain insights into PSC stability via machine learning<sup>2-4</sup>. The recently published Perovskite Database Project<sup>5</sup>, which collected > 42,400 PSCs data extracted from literature, is a major step towards machine learning application in the PSCs field. Despite the database's large size, less than 20% of data points (~7500) have any degradation data available, which is also not necessarily homogeneous. Graniero et al. have shown that the degradation data in this database has low quality to apply supervised machine learning algorithms properly<sup>6</sup>. The study points

out that rather than adding more low-quality data points, a higher data quality (i.e. with more complete information) is needed to train machine learning algorithms. Despite this issue, Zhang et al. recently performed a statistical analysis on the Perovskite Database Project by introducing a new figure of merit for stability called  $T_{S80m}$ . They project the measured  $T_{S80}$  (the time taken to reach 80% of the stabilised efficiency at the end of the burn-in region)<sup>3</sup> to  $T_{S80m}$ , which is the predicted value under reference conditions (300 K, 20% relative humidity, and 1 sun illumination). The authors re-calculate  $T_{S80}$  with the help of acceleration factors which are determined based on various assumptions and which consider the temperature, humidity, and illumination levels during the actual ageing test<sup>7</sup>. While this choice enables a more rigorous statistical analysis, the uncertainty from co-dependencies between different stressors and the range of parameters lowers the accuracy of

<sup>1</sup>Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany. <sup>2</sup>Department of Business Informatics, Freie Universität Berlin, 14195 Berlin, Germany. <sup>3</sup>These authors contributed equally: Noor Titan Putri Hartono, Hans Köbler. ✉e-mail: [titan.hartono@helmholtz-berlin.de](mailto:titan.hartono@helmholtz-berlin.de); [antonio.abate@helmholtz-berlin.de](mailto:antonio.abate@helmholtz-berlin.de)

calculated  $T_{S80m}$  and the analysis does not consider the shapes of degradation curves.

Here, we present a statistical analysis on a highly homogeneous dataset of maximum power point tracking (MPPT) ageing curves, collected during the past 3 years under controlled environmental conditions in a custom-built High-throughput Ageing System<sup>8</sup>. We analyse the power conversion efficiency (PCE) loss after 150 h relative to the maximum efficiency reached during the ageing test. We consider the maximum efficiency to reflect the maximum potential or capability of the solar cell under test, which is why it is chosen as a figure of merit. The dataset comprises MPP-tracks of devices of various structures, including lead-halide perovskite absorbers (both organic and inorganic), charge-selective layers (small molecules, polymer, and inorganic), diverse contact layers (silver, copper, and gold), and architectures (both *n-i-p* and *p-i-n*). See Supplementary Note 1: Data Quality, Supplementary Tables 1 and 2 for a detailed description.

In addition, since PSCs show large variations in degradation curve shapes, there is still no universal metric for PSC stability. Our dataset shows various degradation curve shapes that can be categorised into different groups (see Supplementary Fig. 1). In this study, we perform an unsupervised machine learning method called self-organising map (SOM) to obtain degradation curve shape clusters from the MPP-tracking dataset, where we are able to identify the dominant shape of ageing curves. We observe a correlation between the occurrence of particular shape clusters and the maximum reached PCE. We believe that categorising ageing test data and identifying the main shapes will bring PSC research closer to finding suitable lifetime metrics for PSCs.

## Results and discussion

### Dataset description

We collected 2245 MPPT curves of various device architectures, layers, and perovskite composition materials from August 2019 to August 2022 in the HySPRINT Laboratory at Helmholtz-Zentrum Berlin, Germany. The ageing tests were performed in a custom-built ageing setup<sup>8</sup> under continuous illumination at 1 sun and with individual MPP-tracking for every single solar cell. Variations of the ageing conditions are the device temperature and the use of a UV filter. A summary of the ageing conditions is shown in Supplementary Information Table 1. All devices were fabricated within the same laboratory, albeit fabricated by different researchers (total: ~33 researchers). Within the dataset, 502 solar cells have an *n-i-p* architecture and 1,743 solar cells have a *p-i-n* architecture. The devices have different electron transport materials (e.g. TiO<sub>2</sub>-c, C<sub>60</sub>/BCP, PCBM, and others), hole transport materials (e.g. spiro-OMeTAD, MeO-2PACz, NiO, PTAA, and others), and top electrodes. The perovskite absorber material dominating the dataset is the so-called triple cation (3CAT) perovskite<sup>9</sup> with the general formula Cs<sub>x</sub>MA<sub>y</sub>FA<sub>z</sub>PbI<sub>m</sub>Br<sub>n</sub> (881 cells, with  $x+y+z=1$  and  $m+n=3$ ); other absorbers present in the dataset include CsPbI<sub>3</sub> (218 cells) and FAPbI<sub>3</sub> (56 cells). A complete breakdown of the cell numbers based on the material layers is shown in Supplementary Information Table 2.

### Cell grouping and relative change in PCE

Only the first 150 h of all ageing experiments are analysed to have the maximum amount of comparable data (see Supplementary Fig. 2 for degradation data length distribution in this dataset), and data points which reach maximum efficiency beyond 150 h are excluded (see Supplementary Note: Degradation Time Length for a detailed description). To investigate trends in dependency of the maximum power conversion efficiency (PCE), the ageing tracks are grouped based on the maximum PCE reached during the first 150 h of MPP-tracking as follows: <10.4%, 10.4–14.2%, 14.2–16.8%, 16.8–19.2%, and >19.2%. The data points are divided so that every maximum PCE group has an equal number of cells (449 cells/ group). Maximum PCE is chosen as a parameter of interest because many devices undergo an initial phase of grave losses in PCE at the beginning of the experiment,

also known as the burn-in period<sup>3</sup>. In some cases however, devices undergo an initial gain in efficiency, so-called light soaking improvements<sup>3</sup> and as a result, their initial PCE will not compare well to devices for which the MPP-track immediately decays and for which the initial PCE equals the maximum PCE. Consequently, the maximum reached PCE is selected as value of interest since it reflects the maximum potential or capability of the solar cell under test more universally between different ageing curve shapes. The change of PCE observed after ageing for 150 h relative to the maximum PCE in a cell ( $\Delta PCE, rel$ ) is calculated for each of the different efficiency groups, according to Eq. 1.

$$\text{Relative change in PCE } (\Delta PCE, rel) = \frac{\text{Maximum PCE} - \text{PCE (150 h)}}{\text{Maximum PCE}} \quad (1)$$

This figure of merit reflects the percentage of PCE loss after 150 h concerning the maximum efficiency reached. Note that the relative change  $\Delta PCE, rel$  will include the ‘burn-in’ phase if the MPP-track shows an immediate decay (and the maximum PCE equals the initial PCE), while it excludes the meta-stable phase if the MPP-track shows an initial gain where the maximum is reached after the stabilisation phase. However,  $\Delta PCE, rel$  should be understood as the relative loss observed after 150 h of ageing time with respect to the maximal capability of the system. See Supplementary Note: MPPT Behaviour and Supplementary Fig. 3 for a detailed discussion. The relative change  $\Delta PCE, rel$  is plotted over the efficiency groups in Fig. 1b.

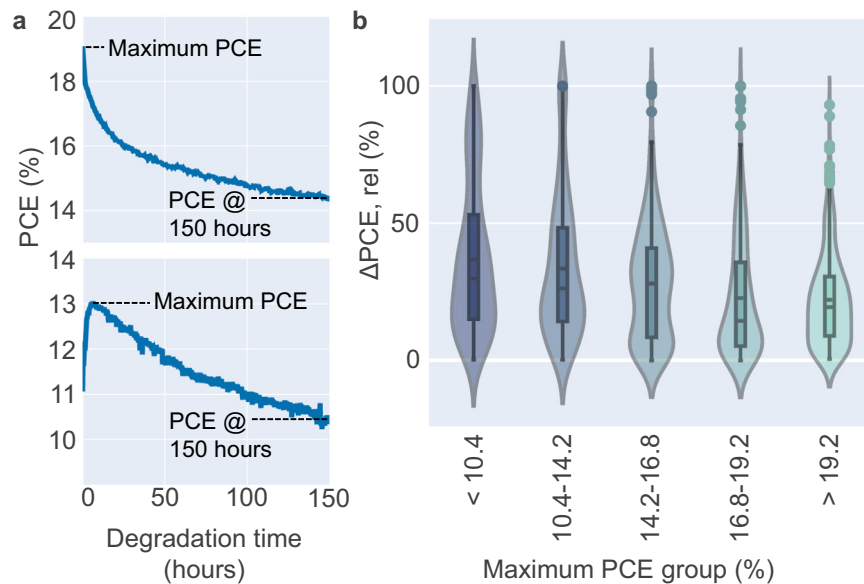
An observation in Fig. 1b shows that as the maximum PCE reached increases, the mean of the  $\Delta PCE, rel$  decreases. This trend suggests that solar cells reaching a higher maximum efficiency during the ageing experiment, statistically also offer a lower loss in efficiency in the first 150 h and can be considered to be more stable according to this figure of merit for stability. The trend is supported by the linear regression performed on the mean of the maximum PCE group and the mean of the  $\Delta PCE, rel$  shown in Fig. 2. Based on the linear regression, for every 1% increase in maximum PCE reached during the first 150 h of degradation, the  $\Delta PCE, rel$  is reduced by -1.5%.

We point out that the observation that more efficient devices also provide higher stability should not be taken literally as a design rule. Specific device layers are known to improve PCE at the expense of long-term stability and vice versa (e.g. inorganic perovskites, spiro-OMeTAD as hole transport layer, and carbon electrodes). The statistical statement given here only illustrates the general trend when looking at the wide range of devices with various stacks and, therefore, different degradation mechanisms. Yet, when looking across different PSCs with various intertwined degradation mechanisms, we see a clear trend that more efficient cells are statistically more likely to have a longer lifetime. Understanding the underlying reasons behind such a trend can strengthen our understanding of PSC ageing behaviour and accelerate the development of highly stable devices.

We propose two hypothetical explanations for the observed relationship between PCE and stability with different causality.

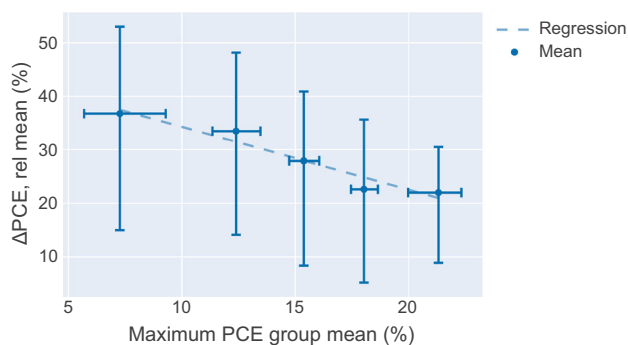
Firstly, at the high level, statistically better stability of high-efficiency PSCs might be explained using a simple conservation of energy model, which states that the total energy of a system stays constant over time. A perovskite solar cell could be modelled as a system that converts incident solar into electric power, as illustrated in Fig. 3. The lower the efficiency of a solar cell, the larger is the amount of energy that remains in the system.

Generally, unavoidable processes lowering the power output of a single junction solar cell are Carnot, Boltzmann and emission losses, thermalisation of above-band-gap photons, and transmission of below-band gap photons<sup>10</sup>. According to the detailed balance limit of efficiency by Shockley–Queisser<sup>11</sup>, this leaves -43–48 % of the input energy



**Fig. 1 | Statistical analysis of a large MPPT-ageing dataset.** **a** Schematic on how the relative change between maximum PCE and PCE at 150 h ( $\Delta PCE, rel$ ) is calculated. **b** The  $\Delta PCE, rel$  for the five groups of maximum PCE. The solid line in the box represents the median, the dashed line in the box represents the mean, the error

bar represents the corresponding interquartile range (between the 25th and 75th percent quartile), and the shaded area represents the distribution of the data points for the group.



**Fig. 2 | Regression fit of the maximum PCE group mean.** The relationship between the mean for each maximum PCE group and its respective mean of the  $\Delta PCE, rel$  is shown. The dashed line is the regression fit of the mean points. The error bar for both axes represents the interquartile range between the 25th and 75th percentile of the maximum PCE group and the  $\Delta PCE, rel$  data.

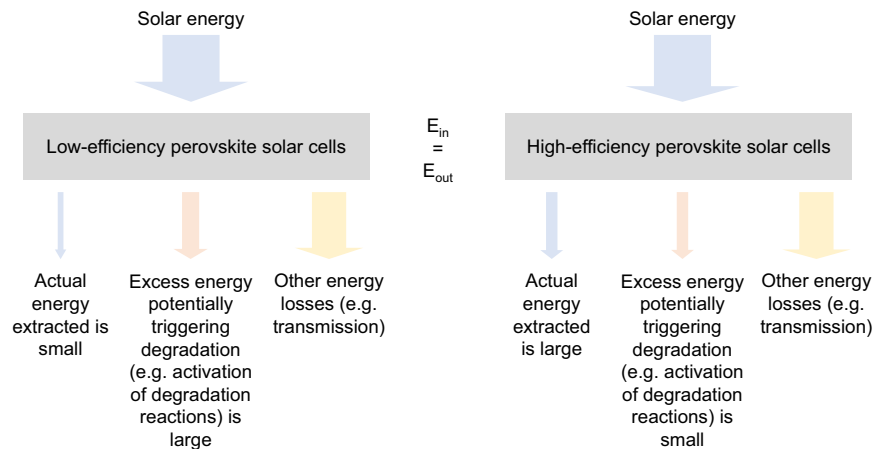
to be absorbed, and a maximum ~30% of the input energy to be converted into electric energy by a solar cell under AM1.5 G for a band gap of 1.6–1.7 eV<sup>9</sup> corresponding to the triple cation perovskite which provides the majority of data points of the present data. Assuming an equal or similar band gap and similar optical properties, the differences between a low and a high efficient solar cell originate in recombination<sup>12</sup> or transport losses<sup>13</sup>. The fraction of absorbed energy, which is not extracted again from the device electrically, remains in the solar cell and gets dissipated in the device where it can potentially trigger degradation. For example, in a 10% efficient “triple cation perovskite (energy absorbed ~466 W/m<sup>2</sup>) device, ~366 W/m<sup>2</sup> of the input energy would be potentially available to cause damage under 1 sun, while for a 20% device it would be ~266 W/m<sup>2</sup>. For an estimation of the influence of the band gap in this dataset<sup>9,14–21</sup> on this theory, see Supplementary Information: Influence of band gaps, Supplementary Table 3, Supplementary Figs. 4 and 5. In the case of trap-assisted recombination, which is likely to be non-radiative, the energy would be transferred to heat that can initiate or accelerate degradation

mechanisms in the device. If the device efficiency is lowered by transport limitations, excess charges will remain in the device and can potentially trigger degradation processes. For example, Lin et al. have shown that the insufficient extraction of generated charge carriers leads to a stimulation of ion migration causing enhanced degradation of devices<sup>22</sup> and Di Girolamo et al. reported that injected charges induce phase separation<sup>23</sup>.

The theory that excess energy might be responsible for instability is also in accordance with the observation that  $V_{MPP}$  is the mildest electronic load ageing condition for PSCs<sup>24</sup> and also other PV technologies<sup>25</sup>. The reasoning is that at MPP, the maximum possible amount of electric energy is extracted from the device, leaving less energy or charges in the device potentially triggering degradation, which leads to the higher stability observed under MPP-tracking compared to  $J_{SC}$  (short-circuit current) and  $V_{OC}$  (open-circuit voltage) conditions. At  $J_{SC}$  and  $V_{OC}$  conditions, no electric energy is extracted from the system and all absorbed energy is potentially available to trigger degradation processes. Note that under  $V_{OC}$  condition additional effects might come to play<sup>25–27</sup>.

We want to point out that our high-level statistical analysis is performed over various PSCs with a wide range of material layers used. Hence it is very likely that also various underlying physical degradation mechanisms are present and it is impossible to directly relate the results of the analysis to the physical causes of degradation. Therefore, the given theory of the conservation of energy should also be understood as a high-level explanation.

A second potential explanation for the correlation between efficiency and stability may lie in the presence of pinholes and defects, incomplete solvent removal, or generally poor device quality. This would put some devices at a disadvantage, and the “fresh device” is already defective at the start of the ageing test, lowering the PCE. Degradation mechanisms could be triggered at defect sites or unremoved solvent. Therefore, the presence of imperfections of any origin might affect efficiency and stability simultaneously. In this potential explanation, the causality is that the same reasons that make a device low in efficiency is also causing a device to be unstable, while in the energy conversion model the low efficiency would make devices unstable as a secondary cause.



**Fig. 3 | The energy conservation schematic for the perovskite solar cell system.** The incident solar power coming into the system equals the power coming out. In the case of high-efficiency cells, the actual power extracted and converted into

electricity is relatively large, leaving a smaller amount of energy to potentially trigger degradation.

### Degradation curve shape clustering

We observe large variations of degradation curve shapes in the first 150 h of the degradation process (see Supplementary Fig. 1). Hence, a machine learning method called self-organising map (SOM) was implemented to find the main degradation curve shapes in the dataset. SOM is a popular unsupervised machine learning method for clustering data by converting the nonlinear statistical relationships of high-dimensional data into geometric relationships in low-dimensional nodes while preserving the topological structure of data, i.e. the relations between one data point and others<sup>28</sup>. In short, SOM compresses information of high-dimensional data, providing visualisation and abstraction of the data.

SOM is utilised to identify the distinct types of degradation curves' shapes based on the normalised PCE data. Four clusters of curve shapes are identified and shown in Fig. 4a: initial gain (cluster 1, total: 1324 points), slow exponential decay (cluster 2, total: 722 points), medium exponential decay (cluster 3, total: 237 points), and fast-exponential decay (cluster 4, total: 97 points). Four is the optimum number of clusters based on the elbow plot analysis shown in Supplementary Fig. 6 and discussed in Supplementary Information: SOM Quantisation Error. In this case, the SOM parameters used explicitly are  $\sigma = 0.5$  and learning rate = 0.1.

While it is known that perovskite solar cells can show very diverse ageing behaviours<sup>3</sup> (see also Supplementary Fig. 1 for a higher number of clusters), this is an impressive example of how diverse ageing tracks are in reality. The vast deviations between the cluster shapes clearly show that the commonly used device lifetime metric  $T_{80}$  (among others) cannot be considered universal for perovskite solar cells. A new figure of merit or a set of parameters that can work across different shapes of degradation curves is therefore urgently needed to compare the ageing results of PSCs.

The normalised count of different clusters for each maximum PCE group is shown in Fig. 4b. As the maximum PCE reached increases, cluster 1, initial gain, increases in share and becomes the majority (> 50% of the share) of degradation curves' shapes for higher maximum PCE groups, showing that more optimum devices tend to have such ageing curves. Simultaneously, the fraction of degradation curves with cluster 4, fast-exponential decay decreases, meaning that high-efficiency devices show less of this type of failure. Those trends can still be observed when the clustering is performed with other possible SOM parameter values (see Supplementary Fig. 7). The clustering results generated by SOM also agree with the result generated with the  $k$ -means clustering method<sup>29</sup> (see Supplementary Fig. 8). Moreover, the slight dominance of cluster types in

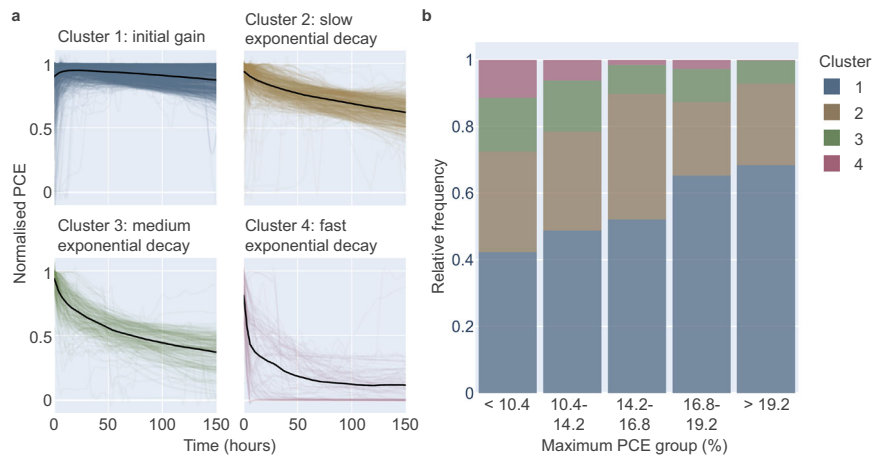
dependency of the architecture (majority of  $p$ - $i$ - $n$  architecture devices have initial gain shape, and majority of  $n$ - $i$ - $p$  devices have exponential decay shapes) are in agreement with an earlier single-stack report by Saliba et al.<sup>30</sup> (see Supplementary Information: Device Architecture Impact on Clustering).

In Fig. 5, the relative change in efficiency  $\Delta PCE_{rel}$  is shown in dependency of the maximum PCE group and sorted by cluster type (a table with the statistics is provided in Supplementary Table 4). The histogram in accord with Fig. 5 is also available in Supplementary Fig. 9. Cluster 4 fast-exponential decay clearly has the largest  $\Delta PCE_{rel}$  with most values located around 100% relative change (i.e. fully degraded). Within the same cluster 4 fast-exponential decay, no data point comes from > 19.2% maximum PCE group, highlighting again that higher efficient devices do not show this catastrophic failure. Defective fresh devices could be one of the reasons that some of the cells, especially from cluster 4, exhibit an instant failure behaviour within the first few hours of the ageing test.

We can also observe that the distribution of  $\Delta PCE_{rel}$  in dependency of maximum PCE shown in Fig. 1b possesses contributions from different clusters and is composed of the combination of those. This is most apparent in the maximum PCE group of <10%, where all types of clusters are represented and a large spread in the distribution is seen in Fig. 1b. A possible outcome of this analysis is that in future ageing tests, the curve shape might be a predictor of stability: If we can observe the initial gain curve shape during the first couple hours, devices have a higher likelihood to be stable within the first 150 h.

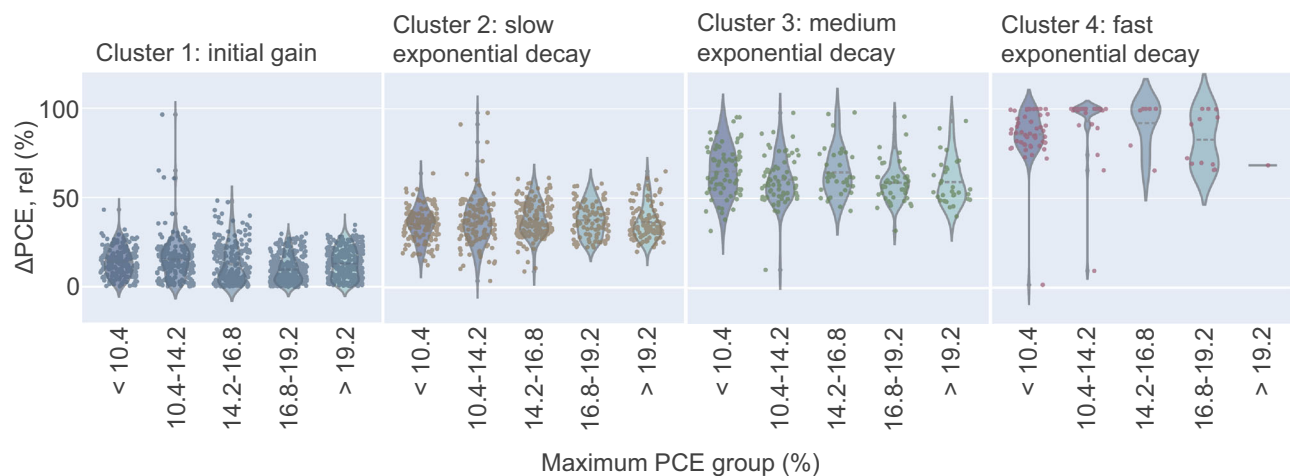
In this study, we analysed a large dataset of MPPT-ageing data collected in-house. The specific aim was to investigate the relationship between the maximum PCE reached during the first 150 h of operational MPPT ageing and the relative loss after 150 h  $\Delta PCE_{rel}$  as a metric for PSCs stability. We discovered that the higher the maximum PCE reached in the first 150 h of testing is, the lower is the mean of  $\Delta PCE_{rel}$ . While this statistical relationship cannot be generalised, it is an encouraging finding that efforts in improving PSC's efficiency go alongside enhancing stability. The finding could be explained using an energy conservation model. Another possible explanation is that increased defect densities render devices to have low efficiency and simultaneously lower stability, as defects might also act as initialisation points for degradation.

Secondly, we clustered the MPPT curves regarding their shape and we discovered that the degradation curve shape is also related to both the cells' maximum PCE group and stability. The initial gain degradation curve shape delivers lower  $\Delta PCE_{rel}$  than other shapes, and this shape type is seen more frequently on cells with higher



**Fig. 4 | Degradation curve shape clustering.** **a** The four different clusters of PSCs' degradation curve shapes categorised through the self-organising map (SOM) method (parameters: sigma = 0.5, learning rate = 0.1). Cluster 1: initial gain, cluster 2: slow exponential decay, cluster 3: medium exponential decay, and cluster 4: fast exponential decay. The black lines represent the centre average of the curves

belonging to each cluster. **b** The relative frequency of each cluster for a given maximum PCE group represented in a bar plot. The violet box represents the fast exponential decay shape (cluster 4), which decreases significantly in occurrence for higher maximum PCE groups, while initial gain shape (cluster 1) can be observed more frequently for more efficient devices.



**Fig. 5 | The distribution of  $\Delta PCE, rel$  based on clusters.** The data points are grouped into previously calculated clusters, and their respective maximum PCE group. In the fast-exponential decay group (cluster 4), there is no data point for the >19.2% maximum PCE group.

maximum PCE. This might act as an early indicator for stability: if the initial gain curve shape is observed during the first couple of hours, devices have a higher likelihood to be stable during the first 150 h.

## Methods

### Ageing of solar cells

Solar cells were aged in a custom-built High-throughput Ageing System<sup>8</sup>. Special electronics were used to MPP-track all cells individually. A perturb and observe algorithm<sup>31</sup> with a delay time of 1 s and a voltage step-width of 0.01 V was applied to track the MPP.  $PCE_{MPP}$  values were taken every 2 min for all cells automatically. Devices were constantly kept at 25, 45, 65 or 85 °C (see Supplementary Table 1) by the use of actively controlled Peltier-elements. Active areas of devices were touching a heat pad for direct thermal coupling. Ageing was performed under a continuous flow of nitrogen (in some cases air) in a closed box, no additional encapsulation was used in the majority of samples (2220). Twenty-five cells were encapsulated with a glass-to-glass encapsulation. Sunlight with 1 sun intensity was provided by a metal-halide lamp using a H6 filter. Supplementary Fig. 10 shows the spectrum of the light source in comparison to AM1.5G. The light intensity was actively

controlled with the help of a silicon irradiance-sensor which was calibrated using a KG3-filtered silicon reference cell certified by Fraunhofer ISE. In 475 cells, a UV filter was used to block UV-light with wavelengths below 380 nm. The ageing test conditions are summarised in Supplementary Table 2 in Supplementary Information. Tests are in accordance with the protocols ISOS-L-II or ISOS-L-2I<sup>3</sup>.

### Data analysis

All the data pre-processing and analysis steps are done in Python. MPPT data pre-processing involves resampling to 10 minutes (since the frequency of measurement varies for each data point), interpolation using the Akima method<sup>32</sup>, normalisation based on the maximum point of the data (MaxAbsScaler) using scikit-learn<sup>33</sup> package, and applying a Savitzky-Golay filter<sup>34</sup> to reduce the noise within the dataset using SciPy<sup>35</sup> package (with window length parameter = 71).

The MaxAbsScaler normalisation is performed on the data by dividing the MPPT PCE across time with the maximum MPPT PCE reached in the first 150 h of ageing test (Eq. 2).

$$x_{\text{MaxAbsScaler}} = \frac{x}{\max(|x|)} \quad (2)$$

The self-organising map<sup>28</sup> analysis is done using the MiniSOM<sup>36</sup> package, and the linear regression and *k*-means clustering is performed using the scikit-learn<sup>33</sup> package. The parameters for each machine learning model are stated under the respective results in the manuscript.

### Data availability

The MPPT data used in this study are available in the Zenodo database under accession code [8185883](https://zenodo.org/record/8185883)<sup>37</sup>.

### Code availability

The Jupyter Notebook with the code for running the analysis is available on Zenodo under accession code [8181602](https://zenodo.org/record/8181602)<sup>38</sup> and [Github](https://github.com).

### References

- National Renewable Energy Laboratory. Best Research-Cell Efficiency Chart | Photovoltaic Research | NREL. <https://www.nrel.gov/pv/cell-efficiency.html> (2022).
- Li, Z. et al. Scalable fabrication of perovskite solar cells. *Nat. Rev. Mater.* **3**, 1–20 (2018).
- Khenkin, M. V. et al. Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures. *Nat. Energy* **5**, 35–49 (2020).
- Sun, S. et al. A data fusion approach to optimize compositional stability of halide perovskites. *Matter* **4**, 1305–1322 (2021).
- Jacobsson, T. J. et al. An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles. *Nat. Energy* **7**, 107–115 (2022).
- Graniero, P. et al. The challenge of studying perovskite solar cells' stability with machine learning. *Front. Energy Res.* **11**, <https://doi.org/10.3389/fenrg.2023.1118654> (2023).
- Zhang, Z., Wang, H., Jacobsson, T. J. & Luo, J. Big data driven perovskite solar cell stability analysis. *Nat. Commun.* **13**, 7639 (2022).
- Köbler, H. et al. High-throughput aging system for parallel maximum power point tracking of perovskite solar cells. *Energy Technol.* **10**, 2200234 (2022).
- Saliba, M. et al. Cesium-containing triple cation perovskite solar cells: improved stability, reproducibility and high efficiency. *Energy Environ. Sci.* **9**, 1989–1997 (2016).
- Hirst, L. C. & Ekins-Daukes, N. J. Fundamental losses in solar cells. *Prog. Photovolt. Res. Appl.* **19**, 286–293 (2011).
- Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys.* **32**, 510–519 (1961).
- Kirchartz, T. & Rau, U. What makes a good solar cell? *Adv. Energy Mater.* **8**, 1703385 (2018).
- Thiesbrummel, J. et al. Universal current losses in perovskite solar cells due to mobile ions. *Adv. Energy Mater.* **11**, 2101447 (2021).
- Møller, C. K. Crystal structure and photoconductivity of caesium plumbohalides. *Nature* **182**, 1436–1436 (1958).
- Emery, Q. et al. Encapsulation and outdoor testing of perovskite solar cells: comparing industrially relevant process with a simplified lab procedure. *ACS Appl. Mater. Interfaces* **14**, 5159–5167 (2022).
- Tsarev, S. et al. A new polytriarylamine derivative for dopant-free high-efficiency perovskite solar cells. *Sustain. Energy Fuels* **3**, 2627–2632 (2019).
- Chen, B. et al. Grain engineering for perovskite/silicon monolithic tandem solar cells with efficiency of 25.4%. *Joule* **3**, 177–190 (2019).
- Chen, W. et al. Surface reconstruction for stable monolithic all-inorganic perovskite/organic tandem solar cells with over 21% efficiency. *Adv. Funct. Mater.* **32**, 2109321 (2022).
- Eperon, G. E. et al. Formamidinium lead trihalide: a broadly tunable perovskite for efficient planar heterojunction solar cells. *Energy Environ. Sci.* **7**, 982–988 (2014).
- Sutton, R. J. et al. Bandgap-tunable cesium lead halide perovskites with high thermal stability for efficient solar cells. *Adv. Energy Mater.* **6**, 1502458 (2016).
- Yao, Q. et al. Dual sub-cells modification enables high-efficiency n-i-p type monolithic perovskite/organic tandem solar cells. *Adv. Funct. Mater.* **33**, 2212599 (2023).
- Lin, Y. et al. Excess charge-carrier induced instability of hybrid perovskites. *Nat. Commun.* **9**, 4981 (2018).
- Di Girolamo, D. et al. Ion migration-induced amorphization and phase segregation as a degradation mechanism in planar perovskite solar cells. *Adv. Energy Mater.* **10**, 2000310 (2020).
- Domanski, K., Alharbi, E. A., Hagfeldt, A., Grätzel, M. & Tress, W. Systematic investigation of the impact of operation conditions on the degradation behaviour of perovskite solar cells. *Nat. Energy* **3**, 61–67 (2018).
- Khenkin, M. V., Anoop, K. M., Katz, E. A. & Visoly-Fisher, I. Bias-dependent degradation of various solar cells: lessons for stability of perovskite photovoltaics. *Energy Environ. Sci.* **12**, 550–558 (2019).
- Prete, M. et al. Bias-dependent dynamics of degradation and recovery in perovskite solar cells. *ACS Appl. Energy Mater.* **4**, 6562–6573 (2021).
- Kim, D. et al. Light- and bias-induced structural variations in metal halide perovskites. *Nat. Commun.* **10**, 444 (2019).
- Kohonen, T. *Self-Organizing Maps*. Vol. 30 (Springer Berlin Heidelberg, 2001).
- Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).
- Saliba, M., Stolterfoht, M., Wolff, C. M., Neher, D. & Abate, A. Measuring aging stability of perovskite solar cells. *Joule* **2**, 1019–1024 (2018).
- Rakocevic, L. et al. Reliable performance comparison of perovskite solar cells using optimized maximum power point tracking. *Sol. RRL* **3**, 1800287 (2019).
- Akima, H. A new method of interpolation and smooth curve fitting based on local procedures. *J. ACM* **17**, 589–602 (1970).
- Pedregosa, F. et al. Scikit-learn: machine learning in Python. *Mach. Learn. PYTHON* **6**.
- Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
- Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
- Vettigli, G. MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map. <https://github.com/JustGlwing/minisom/> (2018).
- Hartono, N. T. P. et al. Perovskite solar cells ageing dataset. <https://doi.org/10.5281/zenodo.8185883> (2023).
- Hartono, N. T. P. noortitan/PvkSOM: first release. <https://doi.org/10.5281/zenodo.8181602> (2023).

### Acknowledgements

H.K. acknowledges the VIPERLAB project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement N°101006715. P.G. acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRIDS). M.K. acknowledges the support of European partnering protect TAPAS (PIE-0015). C.U. acknowledges support by the Helmholtz Association under the programme "Energy System Design". All authors acknowledge the researchers who have shared their PSCs data: Ghafar Adwan, Ece Aktas, Amran Al-Ashouri, Laura Canil, Pietro Caprioglio, Janardan Dagar, Carola Ferber, Emilio Gutierrez-Partida, Juanita Hidalgo, Zafar Iqbal, Jose Jerónimo-Rendón, Ivona Kafedjiska, Eike Köhnen, Bor Li, Guixiang Li, Jinzhao Li, Xin Li, Silvia Mariotti, Natalia Maticiuc, Jorge Pascual, Nga Phung, Manuel Piot, Marcel Roß, Rajarshi Roy, Florian

Scheler, Silver-Hamill Turren-Cruz, Qiong Wang, Fengjiu Yang, Jorge Pascual, Thomas Gries.

### Author contributions

H.K. collected the degradation measurement data, including the cells' parameters and fabrication data and provided devices for ageing. N.T.P.H, H.K., and P.G., performed data cleaning. N.T.P.H. performed data pre-processing and machine learning analysis. M.K., R.S., C.U., and A.A. conceptualised and supervised the research. N.T.P.H, H.K., and A.A. wrote the first version of the manuscript. All authors participated in work planning, reviewing, and editing the manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-40585-3>.

**Correspondence** and requests for materials should be addressed to Noor Titan Putri Hartono or Antonio Abate.

**Peer review information** *Nature Communications* thanks Jingshan Luo, Jiaxue You and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023