Current Opinion in Solid State and Materials Science 19 (2015) 212–226

Contents lists available at ScienceDirect

Current Opinion in Solid State and Materials Science

journal homepage: www.elsevier.com/locate/cossms

Degradation science: Mesoscopic evolution and temporal analytics of photovoltaic energy materials

Roger H. French [a,b,c,d,*], Rudolf Podgornik [b,e], Timothy J. Peshek [a,b,d], Laura S. Bruckman [a,b], Yifan Xu [f,g], Nicholas R. Wheeler [a,c], Abdulkerim Gok [a,b], Yang Hu [a,b], Mohammad A. Hossain [a,h], Devin A. Gordon [a,b], Pei Zhao [a,i], Jiayang Sun [f,g,1], Guo-Qiang Zhang [i,j,1]

[a] Solar Durability and Lifetime Extension Center, Case Western Reserve University, Cleveland, OH 44106, United States
[b] Department of Materials Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, United States
[c] Department of Macromolecular Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, United States
[d] Department of Physics, Case Western Reserve University, Cleveland, OH 44106, United States
[e] Department of Physics, University of Massachusetts, Amherst, Massachusetts 01003, United States
[f] Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, United States
[g] Center for Statistical Research, Computing and Collaboration, Case Western Reserve University, Cleveland, OH 44106, United States
[h] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH 44106, United States
[i] Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106, United States
[j] Center for Clinical Investigation, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, United States

ARTICLE INFO

Article history:
Received 10 December 2014
Accepted 20 December 2014
Available online 20 January 2015

Keywords:
Photovoltaics
Degradation
Mesoscopic evolution
Real-world
Data science
Epidemiology
Reliability
Lifetime

ABSTRACT

Based on recent advances in nanoscience, data science and the availability of massive real-world data-streams, the mesoscopic evolution of mesoscopic energy materials can now be more fully studied. The temporal evolution is vastly complex in time and length scales and is fundamentally challenging to scientific understanding of degradation mechanisms and pathways responsible for energy materials evolution over lifetime. We propose a paradigm shift towards mesoscopic evolution modeling, based on physical and statistical models, that would integrate laboratory studies and real-world massive data-streams into a stress/mechanism/response framework with predictive capabilities. These epidemiological studies encompass the variability in properties that affect performance of material ensembles. Mesoscopic evolution modeling is shown to encompass the heterogeneity of these materials and systems, and enables the discrimination of the fast dynamics of their functional use and the slow and/or rare events of their degradation. We delineate paths forward for degradation science.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Energy materials are essential in our modern world and are expected to have useful lifetimes that extend from 25 to greater than 50 years. The need for long lifetimes and large investments are barriers that new energy producing technologies must be surmounted if they are to provide a substantial proportion of global energy. These challenges were made evident by the Li-ion batteries in the Boeing 787 that were predicted to short circuit only once per 10 million flight hours [1–3]. Two adverse events grounded the whole fleet within four months of introduction, which was three orders of magnitude greater events than estimated. We know much about the synthesis, properties, and function of energy materials, but we do not yet know how to address the fundamental degradation science of energy materials under real-world conditions and time spans.

Since the first large testing of crystalline silicon photovoltaics (PV) module's reliability 40 years ago, we have seen widespread global adoption. The first 5 MW PV power plant, developed in the 1970s as part of the DOE Block Grant Program, was predicted to have a 20 year lifetime. The site power decreased 10× faster than the predicted rate and the failed plant was decommissioned after only 5 years [4]. Degradation-induced failures have been an ongoing characteristic of new and promising PV cells [5,6] and other energy materials even as producers continue to offer 25-year warranties.

The science of degradation of energy materials over time frames longer than >1 Gs (31.7 years) is a fundamental challenge of mesoscale science [7,8] and a transformational opportunity for energy

* Corresponding author at: White 536, 10900 Euclid Ave., Case Western Reserve University, Cleveland, OH 44106, United States. Tel.: +1 302 468 6667.
E-mail address: roger.french@case.edu (R.H. French).
[1] Equal contributing authors.

http://dx.doi.org/10.1016/j.cossms.2014.12.008
1359-0286/© 2015 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

materials, for the following reasons. Realistic degradation studies face the same data and modeling challenges as in medicine, sociology or climate science, where "models and observational data together form an inseparable basis for scientific understanding and prediction" [9,10]. Degradation of energy materials distinguishes itself in that it evolves over long time-frames due to a multitude of distinct, complex, and interacting mechanisms that can lead to a variety of slow and/or rare events that eventually cause failure. There are severe knowledge gaps in identifying, modeling, and *reliably* predicting the mesoscopic evolution that produce degradation, and in *establishing an effective monitoring system* of the evolving process of degradation over the relevant timescales to prevent failures (especially catastrophic failures). It is essential to connect the mechanistic degradation pathways and their temporal evolution at the mesoscale so as to enable the identification of improved and longer-lived energy materials in real life. *Hence, the new degradation science examines degradation of a material or system, guided by real-world or realistic outcomes, whose fundamentals include modeling, monitoring, and prediction of a degradation process, as well as intervention, feature selection, and optimization aimed at improvement of materials and reduction of system failures.* A *new interdisciplinary approach to degradation science calls for the involvement of materials science, physics, chemistry, statistics, computer science, engineering, and energy industries in the investigation of real-world degradation of energy material over their lifetimes.* After reviewing current PV energy materials research approaches (Section 2) we shall illustrate how this should be done (Section 3), what has been done recently (Section 4), and what challenges and new directions lie ahead (Section 5).

1.1. Advances in nanoscience and data science

There have been transformative advances which provide the foundations for a new degradation science. First is the multitude of advances in nanoscience from the scientific community and, from the broader world, advances in computers, communication, and computation.

Since the National Nanotechnology Initiative in 2000 [11], there has been tremendous progress including detailed understanding of science at the nanometer scale and from the femto- to attosecond scale [12,13]. Fundamental advances in glasses [14], nanoscale materials [15–18], interactions [19], fabrication and assembly [20–24], and systems [25–28] have ensued. A detailed understanding of the basic nanoscience that underpins the beneficial function of energy materials provides the *first building block* of degradation science.

Within nanoscience, multi-scale modeling of materials [29] which connects microscopic/atomistic mechanisms with higher level, coarse grained, mesoscopic and even macroscopic models [30–34] helps us understand the fundamental origins of the *physical* properties of materials. This multi-scale modeling research has focused on spanning length scales underpinning the effective parameter passing approach, linking atomic scale behavior to experimentally determined macroscopic properties. The Materials Genome Initiative [35,36] is an example of accelerating materials discovery modeling at the micro- and mesoscopic levels to guide the experimental synthesis of new materials [37].

Still, multi-scale modeling has not elucidated the range from femto- to gigaseconds needed to provide fundamental guidance on the mesoscopic evolution of materials and their long term degradation in function and properties. For this challenge, we start with newly available data from large and diverse experiments. These varied datasets provide important information to fit appropriate *physical* and *statistical* models and identify the fundamental mechanisms of energy material degradation. These are then incorporated into a network model of their mesoscale evolution over lifetime. As nanoscience and multi-scale modeling advance, they will continue to provide mutual benefit by illuminating fundamentals and identifying critical contributors and effects.

The tremendous advances in computation and communications [38] and open access, code and data manipulation [39–45] over the past ten years are a *second building block* for the opportunities in degradation science. Distributed computing [46–48] improved Internet connectivity and mobility. The ubiquity of sensors make for unprecedented big data streams which can be utilized for experimental studies of energy materials in the laboratory and in real-world conditions [49]. Data science has grown beyond the purely computational advances which have been the focus of science (e.g., high performance computing). With increased data volumes and variety and the associated advances in informatics for petabyte scale analysis, it is now possible to study large populations of real (as opposed to idealized or simplified) energy materials under real-world conditions and over very long time frames. These epidemiological or population-based studies can complement our traditional small sample size, laboratory-based experiments, providing additional statistically sound information to bridge the 24 orders of magnitude in time required for femto- to gigasecond science.

1.2. Temporal evolution of mesoscopic energy materials

The Materials Genome Initiative and advances in nanoscience have allowed nascent energy materials to be developed; however, a predictive framework for those materials properties over time in real-world applications is lacking. For example, there is much research on new batteries and improved storage capacity for applications in electronics, transportation, and grid [50], yet degradation science must be applied to understand the contributing factors that limit the number of charge cycles and the basic mechanisms and pathways that lead to end-of-life failures [17].

Functional energy materials are complex materials with homo- and heterogeneous interfaces and substantial variances among samples in a population. By virtue of their energy function, they are non-equilibrium systems with cyclic operating conditions and stressors and thus have high spatio-temporal complexity. For energy applications spanning the time domain from femto- to gigaseconds, a new approach is needed that can distinguish the large dynamics of function from the slow/rare events of degradation and their differing temporal regimes. For example, damage initiation, accumulation, and growth will eventually lead to a transition such as a sudden precipitation or a possible bifurcation into a new regime. Similarly, environmental conditions, as encountered in permafrost or desert or given by daily or annual cycles of the seasons, can produce results quantitatively different from a well controlled laboratory-based study. To understand the degradation significance of each of the heterogeneous aspects of materials and devices across populations and characteristic time scales, all respective data needs to be accessible to scientific inquiry. The focus of this research is the development of mesoscopic-evolution network models which integrate *physical* and *statistical* models phenomena. These network models, exemplified in Fig. 1 for poly(-methyl methacrylate) (PMMA) acrylic, link micro- and mesoscopic degradation in order to understand the stressor/mechanism-mode/responses of the PVs in real-world use over their lifetime.

2. Current PV energy materials research approaches

Three distinct communities (scientist, engineers, owners/operators) with distinct goals have worked in research and development of energy materials. Scientists typically pursue laboratory-based research topics related to materials performance. Engineers seek

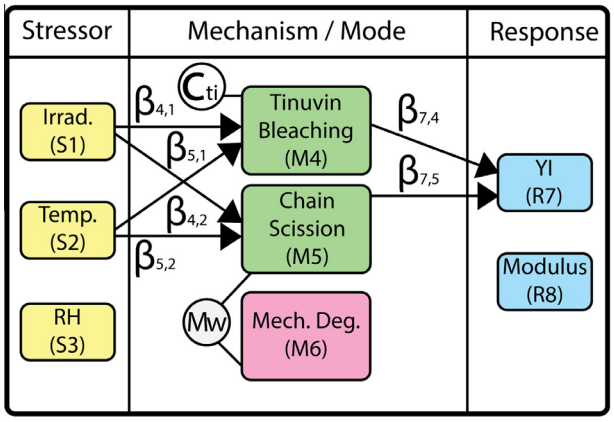

Fig. 1. Network model for the stressor/mechanism-mode/response for acrylic degradation, which shows the contributions of each stressor on a measured response where C_{ti} (concentration of tinuvin) and Mw (molecular weight) are latent variables shown is circles. β_{ij} is the coupling strength. Yellow boxes are stressors. Green boxes are mechanisms. Pink boxes are modes and blue boxes are responses.

standards to guide product development and safety of holistic systems. Owners and operators are most concerned with real-world performance and return on investment. These communities overlap and interact, but their varied approaches have hindered the development and permeation of a cohesive scientific foundation and culture across the energy materials community as a whole.

2.1. Laboratory-based studies

Laboratory-based studies of degradation in PV energy materials have typically focused on a specific observed failure mode and sought to accelerate the presentation of that mode. This methodology has been used in potential-induced degradation [51], however, no *physical* modeling of causes is provided. Novoa et al. [52] examined the bonding strength of PV module backsheets to comparable materials in a simulated PV module in laboratory accelerated aging conditions. Peike et al. [53] studied the effect of the system interactions on the power performance of a solar cell when laminated in various structures. Koehl et al. [54] tied the real-world climatic data of moisture content and hydrolysis reaction rates in data-driven modeling of accelerated aging conditions. *Physical* models of moisture ingress and diffusion into real materials and systems have been employed to estimate water concentration [55]. These studies tended to focus on distinct modes observed in the accelerated aging of PV energy material expected from industry testing standards; however, there is little correlation with real-world observations and few insights that are extrapolated to actual use-conditions.

2.2. Standards-based testing

Internationally-recognized standards for certifying the performance and safety of commercially available PV products are in existence for both technologies [56] that include measurements [57] and accelerated exposure conditions [58]. However, several high-profile fires and other non-conformances have been well-documented [59]. Standardized laboratory-based testing is based on short duration, usually 3–4 months, with binary (pass/fail) outcomes. These tests are designed to assure a uniformity of quality at manufacture, although the tests are widely believed to provide a rapid means of detection for known failures or degradation modes of a product in its intended environment. This statement is highly relevant to this discussion because the failure and degradation modes must be known fully (i.e., there is no predictive quality to these tests and yet there is an implicit linkage to the end-use environment). An example of these test procedures for crystalline silicon PV is IEC 61215 [56]. Due to the binary nature of the results, problems identified by tests cannot include corrective action insights and are subsequently learning limited.

2.3. Real-world studies

An epidemiological approach to data collection and analytics has only recently been applied in real-world studies of PV systems [60–63,5]. The published real-world studies were usually used to monitor and model performance at coarse time scales or to determine average degradation rates over long periods. In these longitudinal studies, as little as one data point per month was collected over time and a simple linear degradation trend was fitted and compiled for sites around the globe. Jordan and Kurtz [64] reported real-world performance of over 2000 power plants to obtain a mean linear model of PV module degradation rate equal to 0.5%/year and compared these results to other studies. However, a set of modules exposed to hot/dry conditions underwent mean degradation rates of 1.5%/year [65]. These approaches are quite different from the approaches used in the climate science community, where statistical analysis is applied to all variables of the system [66] and statistically significant relationships are identified [67].

3. Degradation science: mesoscopic evolution modeling

The basic science of degradation will span the time domains from mechanisms to lifetime and will provide an integrated mesoscopic evolution modeling methodology. This methodology encompasses the active mechanisms and determinants of the performance, safety, and function of energy materials over diverse real-world conditions. The challenge of establishing degradation science as the foundation of energy materials research is both cultural and scientific. The integration of laboratory-based and real-world studies is both a prerequisite for developing a full mesoscopic evolution science and a necessary step for deeper understanding of the effects of degradation. Bringing the approaches of different cultures and communities (scientists, engineers, and owners/operators of essential energy materials and infrastructure) together will be critical to speeding up the invention, implementation, and penetration of new energy materials in the real world beginning with *data acquisition* to the final realization of *reliable prediction and monitoring of degradation.*

The data acquisition strategies must balance the relevant scales and volumes of the datasets to be used in the *physical* and *statistical* modeling. Approaches for extraction of the necessary information must be able to disregard spurious information, so as to develop a working network of models for each active mechanism related to each degradation pathway on the mesoscopic *physical* level and the data-driven *statistical* model level. To capture the temporal evolution of the energy material over long time frames, appropriate informatics methods are needed to balance data volume (e.g., simple univariate time-series data streams with high-dimensional volumetric imaging datasets) while considering their respective information contents [68,69]. The raw data and extracted information must be accessible for query and modeling. Similarly, the modeling approaches used to understand and parameterize active mechanisms and phenomena over lifetime fall into the broad categories of micro-, meso- and macroscopic approaches. Laboratory and real-world experimentation, informatics, analytics, and the development of network models for mesoscopic evolution of energy materials over lifetime together constitute the field of degradation science.

3.1. Integrating laboratory and real-world studies

Laboratory-based experiments seek the deterministic modeling of mesoscale evolution by studying the effects of multiple stressors, experimentally controlled mechanisms, expected active mechanisms, and multiple responses critical to lifetime performance. The stressors of the real world cannot be experimentally controlled, but must be actively monitored and their cumulative effect over time explicitly integrated, as is done in our global SunFarm Network (Fig. 2). In these observational real-world studies, active degradation may include unexpected yet significant mechanisms. The systems under study typically have strong heterogeneity with large variances that increase over time among diverse evolutionary paths and modeled often using stochastic assumptions [70].

When planning laboratory experiments, the number and types of repetition can be determined or estimated using results from a pilot study or similar historical studies to reduce over- or under-sampling and ensure sufficient statistical power [71,72]. While a prospective real-world population-based observational study involves uncontrolled environmental stressors, and special attention is needed to faithfully record potential relevant variables for follow up studies. This is especially important in order to infer causal relationships between stressors, mechanisms, and responses [73,74].

3.2. Stress/mechanism-mode/response framework

It is beneficial to formulate the degradation science mesoscopic evolution models using a stress/mechanism-mode/response framework [75] to target our full library of degradation mechanisms and pathways with sufficient sampling and data acquisition for both real-world and laboratory-based experiments. Thorough planning of the sample populations, evaluations, and datastreams is necessary to maximize the information yield [71,74]. Efficient statistical planning is important to allow the study to minimize bias and ensure reproducibility. To explore or test microscopic mechanisms of response to a specific applied stressor or stressors, which were either observed in a real-world study or hypothesized based on physics theory, carefully designed laboratory-based experiments are essential. Prospective randomized comparative studies repeated over time under various types of stressor conditions (that mimic the real-world environment) will be more effective than retrospective studies in evaluating causal effects of different stressor conditions on different material types. To elucidate the library of possible degradation mechanisms, not only are materials science and physics needed, but also statistics in modeling degradation and identifying important features/factors that impact degradation. The interaction of these sciences from pilot and intermediate studies will advance the degradation science. Ultimately the material degradation library comprises a network of submodels, both *physical* and *statistical*, generating a system of multivariate equations for response prediction given applied multifactor, sequential, or cyclic stressors.

3.3. Informatics and ontology for data integration

We envision an informatics and data analytics environment to manage and support the entire data lifecycle including data assembly, data reuse, query, exploratory data analysis, and data sharing. A pilot environment called E-CRADLE™ (Energy Common Research Analytics and Data Lifecycle Environment) is under development for such a purpose (Fig. 3). The key architecture component of E-CRADLE™ is a domain-specific ontology governing data acquisition and ingestion, data annotation and curation, data assembly, and user-interfaces for accessing data. E-CRADLE™ is designed to help ease the data management burden for the entire community by sharing the expertise and data resources, as well as performing exploratory data analysis for development of *physical* and *statistical* models. A community-wide resource such as E-CRADLE™ would enable pilot studies from a wide range of investigators and would support continued large-scale studies as new data becomes available. Our approach is consistent with and leads in the direction the federal "open science" mandates [39,41,42].

3.4. Physical and statistical submodels

3.4.1. Physical modeling

Physical models usually assume upwards progression in the space–time scales or in the nature of implied coarse graining. The latter is unavoidable since a complete atomistic/quantum description on a microscopic level is usually not realistic and

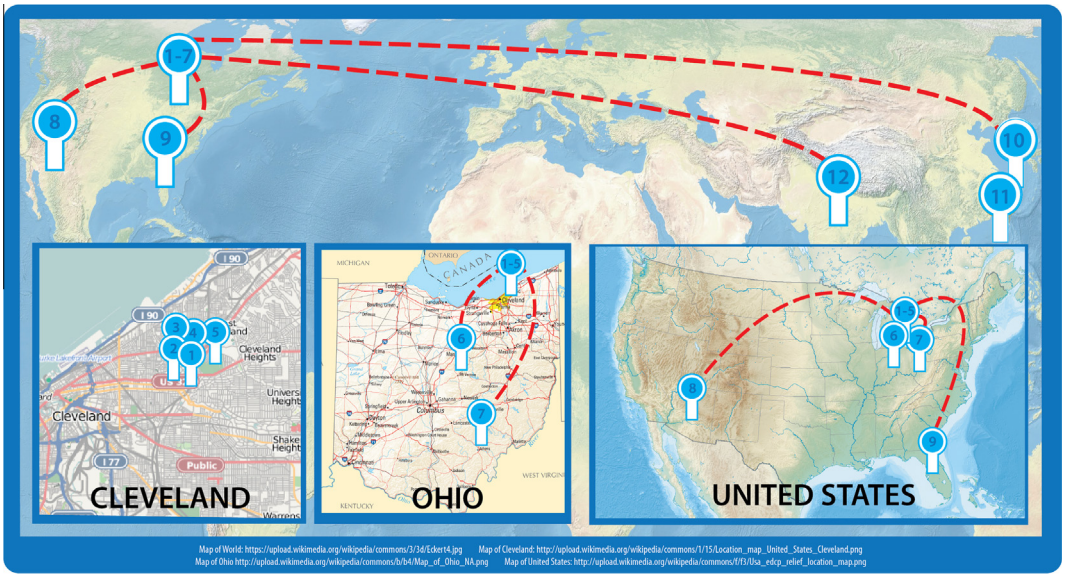

Fig. 2. Real-world data are collected from the global SunFarm network including 12 locations around the world: 1. SDLE Center SunFarm; 2. Lakeview 1 MW power plant; 3. Tinkham Veale Student Center roof; 4. Seidman Cancer Center roof; 5. 1–2-1 fitness roof; 6. Replex SunFarm; 7. AEP Dolan Test Center; 8. Arizona Q-Lab SunFarm; 9. Florida Q-Lab SunFarm; 10. Underwriters Laboratory (UL) Taiwan Lujhu SunFarm; 11. UL Taiwan Taitung SunFarm; and 12. Indian Institute of Technology Gandhinagar, Ahmedabad (IITGN) SunFarm.

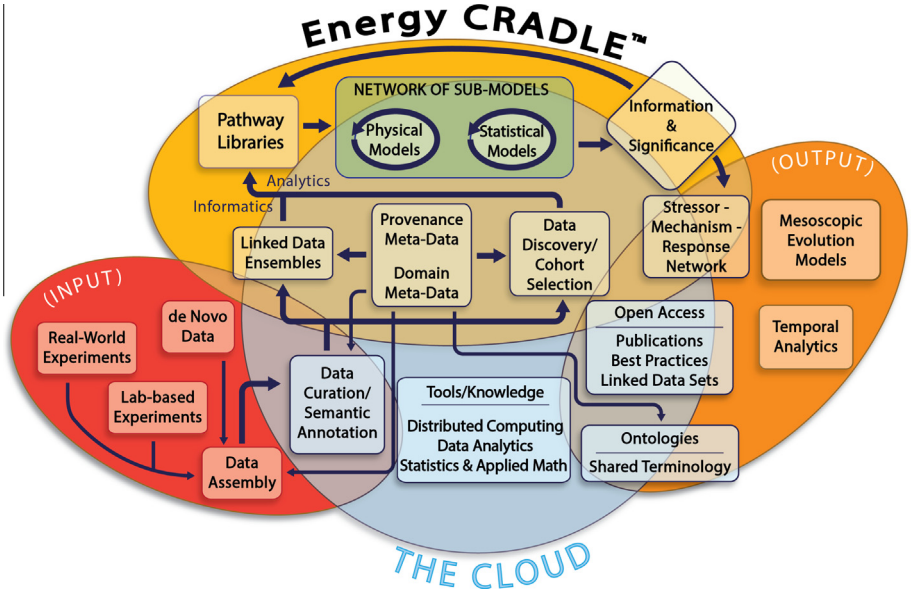

Fig. 3. A schematic of the E-CRADLE™ infrastructure based on distributed cloud computing, data analytics, statistics and applied math. Real-world data, laboratory-based data, and reused data are ingested into the cloud, reassembled under the guide of domain knowledge. With the help of pathway libraries, mesoscopic evolution models of materials are extracted from the network of physical and statistical submodels built on ingested data.

mesoscopic or even macroscopic *physical* models are unavoidable. The range of applicability of these models is limited by the extent of coarse graining that is implied by the experimentally determinable parameters and relations, as shown below:

(a) Microscopic *physical* models specifically refer to atomic/molecular aggregate mechanisms describable by quantum-electrodynamic Hamiltonians as functions of quantum variables solved at a finite temperature. They have in principle no implied coarse graining and no free parameters. Examples include quantum theory of the photovoltaic effect starting from a microstructural lattice-electron description of a semiconductor device or density functional theory of optical spectra of materials [76].

(b) Mesoscopic *physical* models imply first level of coarse graining described with Hamiltonians that do not depend anymore on quantum variables but on coarse grained order parameters with assumed symmetries, coupled with phenomenological constants that only in principle depend on microstructural parameters [77]. Examples include continuum description of various ordering phenomena (liquid–solid phase transition) with Landau-type order parameter theories [78] or the continuum description of structural defects (dislocations and disclinations in crystals) in mesoscopic ordering [79]. These models can seldom be reduced to some features of the microscopic description.

(c) Macroscopic *physical* models depend on macroscopic variables and their connection with the micro-world of quantum variables or even the meso-world of order parameters is very tenuous. They imply even more coarse graining, losing completely the connection with micro- or mesostructural variables and are more based on experiments and statistical data then on micro- or mesoscopic considerations. Examples include Navier–Stokes hydrodynamics [80], reaction–diffusion systems [81], flocking phenomena in birds [82], or epidemiological models in population dynamics [83]. These models are not even in principle reducible to meso- or microscopic underlying descriptions.

Defects and disorder are concepts that arise naturally with growing coarse graining of the meso- and/or macrostructural model description [84]. While thermal disorder leading to thermal fluctuations (or noise) and entropy variation is ever present, structural defects and disorder are usually imposed by external constraints or non-equilibrium driving mechanisms. Specifically, in the context of aging and degradation mechanisms of energy materials, externally controlled or uncontrolled stressors introduce structural defects that can either couple to the thermal fluctuations (annealed disorder) or remain independent of the thermal fluctuations (quenched disorder) [85]. Both, however, introduce degradation mechanisms that lead to a progressive disordering of the material eventually completely destroying its functionality. Examples of the quenched disorder model include photo-oxidation yellowing of the PV energy materials, where defects are chemical and essentially immobile in nature, progressively accumulating under stressor conditions leading to degradation irrespective of the thermal fluctuations. Hazing of polymeric materials would be an example of annealed disorder, where stressor generated microscopic defects interact with the matrix taking advantage of the underlying thermal fluctuations to migrate and nucleate into macroscopic disorder degrading the material's functionality.

3.4.2. Statistical modeling

Statistics is the science of data. Evidence from the real world and implications from experimental studies lie in data. Any statistical analysis of data is based in the model (implicitly or explicitly) and analysis technique (or analytics) used. The model and analytics must be suitable and responsive to the degradation science objectives. The *statistical* models and analytics needed for modern degradation science include, but are not limited to, the following: (a) *Statistical* models for the performance of microinverters and PV modules in real-world and suitable experimental settings, when no physical models exist for guidance. These models may be parametric, semi-parametric, nonparametric, longitudinal, or include change points, depending on data and objectives. (b) Statistical feature selection techniques to find important factors that affect power performance. (c) Statistical measures that handle large and high-dimensional data. (d) Statistical design strategies that structure studies for significance/information extraction, such as those that will lead to and validate accelerated experiments in labs for long-term reliability studies (in cooperation with materials

science). (e) Statistical analysis of observational studies at different depths and (sub) population levels from different locations under different stressors, leading to individualized predictive models that can be updated automatically as more data becomes available, much like how typical weather forecasting is updated from a month ago to a day ago. (f) Statistical process surveillance that monitors and alerts a potential failure or problem of a degradation system. (g) Statistical study of pilot data or interim analyses for updated exploration and discoveries. (h) Statistical methodology for (d), (e), and (f).

In addition to building statistically informed submodels following above steps, a network model (or a system of structural equation models) which links both physical and *statistical* submodels needs to be developed to reveal larger frameworks of degradation systems (Section 3.5).

3.5. Mesoscopic evolution models: networks of submodels

Linking *physical* and *statistical* models into networks of models on different spatial/temporal scales, using torrential real-time data streams, is among the major challenges facing predictive capabilities of the degradation science. In this respect there exists definite similarities with modeling of climate processes [10] and modeling of social evolution processes [86]. However, there is an extra layer of challenge to materials degradation science; specifically, it would take at least 25 years to have real-world reliability data evidence if there was not an effective interaction of physical and statistical models to link the real-word and laboratory studies. Even multiscale modeling, including recent coarse-grained simulations of biophysical and chemical systems [32,87,88], cannot be seen as a guiding paradigm for material degradation because it is driven exclusively by *physical* models with well defined and controlled stressors, leading to the effective parameter passing approach to link the microscale with macroscale properties. Thus, it is imperative to have an evolution of updated data (and studies of these data) to mimic real-world settings as closely as possible (based on physics and statistics) in labs where an accelerated exposure can be made. In this way, it is possible to derive more reliable predictive models of degradation (than existing models) in a much shorter time frame (than 25 years), using a suite of real-world and laboratory data (such as data from our partner SunFarms).

Mesoscopic generalized structural equation models, semi-supervised by domain knowledge (semi-gSEM) [89], based on SEM theory of statistics, epidemiology and mathematical sociology, is a model sample that connects the *physical* and *statistical* mechanistic submodels into a network semi-gSEM model of degradation pathways that encompasses the lifetime temporal evolution. In this way we take simultaneous advantage of the SEMs exploratory and confirmatory *statistical* models, elucidating statistically significant relationships in complex systems with both measured and latent variables, as well as of the intermittent micro-, meso- and macroscopic models with internal mechanistic variables, whenever they are applicable or indeed existent, capturing relationships/couplings among all the variables available from laboratory-based and real-world experiments.

Another important feature of torrential real-time data streams (over extremely long time-scales coupled with flexible networks of embedded submodels) is that they make it possible to detect/analyze the impact of highly improbable events [90], that would be missed in any laboratory-based time framework, and ways the system can self-organize and cope/fail under externally imposed disorder stressors. Recent advances show that there exist materials that show an inherent antifragility [91] in their response to degradation disorder, including solid-electrolyte interfaces [92] and synthetic polymers [93].

4. Degradation science approach exemplified

To understand the degradation science, progress in informatics and statistics developments are made, together with clusters of ongoing laboratory and real-world studies guided by (updated) domain knowledge and updated evidence, especially in the methodology relating stressors with a response in a material or system.

4.1. Informatics: E-CRADLE™ data lifecycle

In order to enable the studies of mesoscopic evolution using temporal analytics, researchers in different fields need an infrastructure encompassing the laboratory-based controlled stressors and the real-world uncontrolled stressors for data sharing and analysis. CRADLE™ serves as the platform of data acquisition, data storage, data processing and data presentation, enabling *physical* and *statistical* model development and application with a focus on mesoscopic evolution and temporal analytics.

E-CRADLE™, based on the multi-modality, multi-resource, information integration environment [94], is the first example of CRADLE™ infrastructure focused on handling the PV system and environmental monitoring data time-series data-stream from the SunFarm network and laboratory-based PV data. All the SunFarms are equipped with minute-by-minute PV power and environmental data monitoring instruments, with live data-streams flowing to the SDLE center [95]. Laboratory-based data are typically spectra and images. About 200 GB of data is generated each year. The nature of the data indicates that it is subject to typical big data problems: high volume, velocity, and variety.

Compared with the traditional way of using MySQL database to store and retrieve data, E-CRADLE™ leverages Hadoop [96], Hadoop distributed file system (HDFS) and HBase [97] and shows significantly better scalability on data processing (up to 20 times faster), data storage and data retrieval [98]. E-CRADLE™ (Fig. 3) is based on Cloudera Express [99] as a combination of CDH and Cloudera Manager. By leveraging Cloudera Express, E-CRADLE™ provides batch processing, interactive query, interactive search, and interfaces with multiple programming language, such as Hive or Pig to process data. In-place analytics on Hadoop is being implemented using RHadoop [100] and other R packages to enable researchers to manage, analyze and model degradation science problems in place without requiring data downloading, thereby enabling a data-centric approach of moving the analytics to the data.

4.2. Mesoscopic evolution modeling: semi-gSEM

We developed a domain-semi-supervised generalized structural equation modeling (semi-gSEM) methodology that can be used to relate *physical* mechanistic submodels and data-driven *statistical* submodels as networks of mechanisms/modes with *statistically* significant pathway relationships [89]. A component of SEM is to have network relationships and couplings amongst variables (stressor/mechanism-mode/response) which can be rank-ordered as contributors for degradation. Temporal evolution, damage accumulation and transitions (i.e. change points) among mechanisms and modes and will be accounted for in these semi-gSEM models which can be combined into a mesoscopic evolution network model.

Our semi-gSEM is similar to the SEM [101,102] from sociology [103], psychology [104], marketing [105], epidemiology [83], or chemical or biological reaction network modeling [106,34] in that there is a system of equations depicting multiple paths of different factors (in our case, the stressors) impacting on intermediate factors (or outcomes) and then the final outcomes (in our case, the

power performance). Our semi-gSEM was generalized to allow different paths to have different functional forms and include nonlinear relationships; and was semi-supervised based on the current knowledge of physics and chemistry relating to degradation mechanisms/modes (e.g., using plausible functional forms such as linear, quadratic, exponential, or logarithmic, along with inclusion of change points). Generalizing the SEM models allows for the different stages of degradation to be effectively modeled.

For small scale networks with simple relations, the resulting semi-gSEM model contains coupling coefficients (β_{ij}) amongst variable pairs (i,j) [89]. The coupling β_{ij}denotes the coefficient vector of the regression model that predicts variable i and j, where the functional form is chosen from a domain guided pool of candidates and the final form is determined by the empirical evidence. β_{ij} can also depend on other intermediate variables such as *Irrad* (S1), *Temp* (S2), and *RH* (S3) that denote temperature, irradiance level and relative humidity among the most influential factors in degradation. The β_{ij} may also subject to an additive error or uncertainty that is due to random error and any unobservable factors such as latent variables that characterize critically important physical/chemical mechanisms but are not directly observed.

A large-scale network model can be conceptually divided into a network among submodels. For example, a study of degradation of acrylic [107] (Fig. 1) includes two submodels, Tinuvin bleaching and chain scission, each forms a degradation mechanism. The concentration of the Tinuvin, denoted by C_{ti}, reduces over time (Tinuvin bleaching). C_{ti} can has significant impact to the system performance, but is a latent variable (i.e., not directly observed). In order to measure the change of C_{ti} and assess its effects on the rest of the system, the physical Tinuvin bleaching (M4) submodel suggests that three *observable* variables, UV absorbance, mechanical degradation, and the rate of Tinuvin bleaching capture the necessary information. The mathematical constraints upon these three variables are also given by the Tinuvin bleaching submodel. Tinuvin bleaching submodel provides a mathematical constraint on above observable variables. Similarly, the chain scission (M5) submodel also provides suitable observable intermediate variables and constraints. The strengths of relations among submodels are denoted by β's, coupling strengths, which denote vectors of coupling coefficients among key variables of stressors, mechanisms/modes, and system response. In this type of network model, we use $\vec{\beta}$, called coupling strength, to denote the vector of coupling coefficients between observable variables of different submodels, stressors and system response.

The current stage semi-gSEM methodology follows two principles. Principle 1 determines the univariate relationships between stressor/mechanism-mode/response variables under the assumed Markovian property that if given the value of the current variable, then future and past variables are independent to each other, or the current variable is sufficient to relate the next level variable. Principle 2 simultaneously considers the stressor variables and the mechanism/mode variables acting on the response variable collectively by an additive model. The final collective additive model is determined by a generalized stepwise variable selection [89].

4.3. Laboratory: photodegradation of acrylic

Acrylic polymer, or poly(methyl methacrylate) (PMMA), has been aged at the SDLE Center according to the degradation science methodology: real-world and accelerated aging studies have yielded data that are cross correlated and a predictive model was built in the semi-gSEM framework. Acrylic is used in PV energy materials that reflect and concentrate light (such as back-surface mirrors) and Fresnel lenses (such as concentrating PV systems [108]). The material must withstand harsh outdoor conditions including UV radiation and other stressors [75], yet unstabilized acrylic is susceptible to photo-, thermal and chemical degradation that manifest as changes in optical and mechanical properties [109]. Stabilizer concentration and degradation in relation to polymer degradation was the subject of a laboratory-based study to elucidate the multivariate system degradation library. A network of submodels, shown in Fig. 1, was developed based upon semi-gSEM methodology to predict lifetime and optimal stabilizer concentrations [107].

Photodegradation studies of both unstabilized acrylic materials and those compounded with Tinuvin UV absorber have shown reciprocity in response to identical doses across a range of irradiance levels (1×, 5×, 50×), confirming that the active degradation mechanisms remain constant independent of irradiance intensity. However, spectral effects are observed when contrasting full spectrum and UV-only irradiance sources – similar active degradation mechanisms are observed in the UV region response (changes in the fundamental absorption edge due to chain scission and UV absorber bleaching), but yellowing progresses much faster in response to the full spectral exposure due to additional photodarkening of the yellowing chromophores.

Data-driven semi-gSEM modeling of these study results give insights into the interrelated mechanisms at play such as changes in the fundamental absorption edge due to chain scission are strongly associated with the yellowing response. In Fig. 4, a mesoscopic evolution model including UV absorber (Tinuvin) bleaching (TB), chain scission (CS), and mechanical degradation (Mech Deg) is shown, which encompasses the three stages in this materials temporal evolution to failure, over a 25-year period. Stage one encompasses tinuvin bleaching until concentration of tinuvin (C_{ti}) reaches zero. A change point is seen in this grade at 12 years (C_{ti}) when mechanical degradation accelerates (stage 2) and is the factor that defines the usable lifetime of this grade of acrylic. This photodarkening effect initially decreases tenfold with the addition of the Tinuvin stabilizer present in multipurpose (MP) grade, with the fundamental mechanism of photobleaching of the Tinuvin UV absorber determining the Tinuvin consumption rate. Once the Tinuvin is consumed, backbone chain scission dominates the degradation (Stage 2), and this is followed by dramatic loss of mechanical properties (modulus) in the third stage of degradation [110].

4.4. Laboratory: photodegradation and hydrolysis of polyester

Similarly to Section 4.2, bulk polymer of polyethylene-terephthalate (PET) samples were exposed to real-world and accelerated aging to model real material degradation under hydrolytic and photolytic mechanisms leading to a transition and loss of mechanical properties. PET is a critical component in photovoltaic backsheets due to its high dielectric strength, but the material is highly susceptible to environmental stresses. PET degradation mainly occurs via photolytic, hydrolytic, and thermal cleavage of an ester bond and results in discoloration and/or hazing, decreased molecular weight, and increased crystallinity. Photodegradation and/or photo-oxidation mainly proceeds via Norrish type I or Norrish type II reactions that determine further degradation pathways [111–114]. Hydrolysis mechanisms are more complex when chemical reactions like autocatalysis due to active carboxylic acid end groups are taken into account and various kinetic models have been discussed in the literature [115–121]. Failures such as cracking and delamination in backsheet films caused by aging [122,52] may result in dielectric withstand breakdown of PV systems.

To study photodegradation of unstabilized PET, a laboratory-based experiment utilizes a completely randomized longitudinal design (i.e., unstabilized PET samples are randomly assigned four exposure types and followed over time with repeated measurements). The four different laboratory-based accelerated conditions

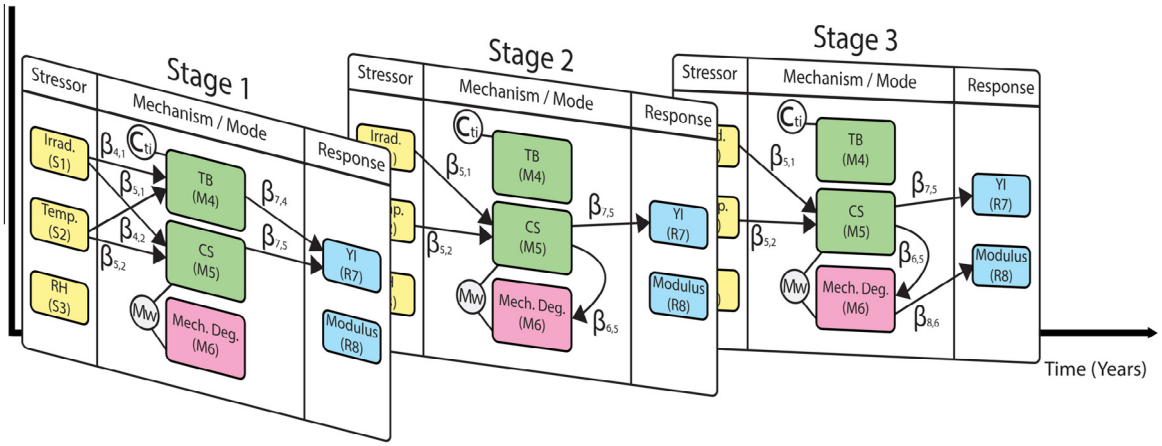

Fig. 4. An illustrative mesoscopic evolution model of degradation pathways in stabilized PMMA acrylic spanning a 25-year lifetime, showing change points among multiple degradation mechanisms and modes. Temporal analytics can be combined with real-world and laboratory experiments to produce the network of submodels. Stage 1 relates to tinuvin bleaching as the largest degradation mechanism contributer. After a change point, Stage 2 is dominated by chain scisson which leads to mechanical degradation in stage 3. TB represents tinuvin bleaching, CS represents chain scission, RH represents relative humidity, C_{ti} represents concentration of tinuvin, Mw represents molecular weight loss, YI represents yellowing index, and β_{ij} represents coupling strengths. Yellow boxes indicate stressors, green boxes indicate mechanisms, pink boxes indicate modes, white circles represent latent variables, and blue boxes indicate responses.

are continuous and cyclical UVA light, heat and humidity in accordance with ASTM G154 [58], and also damp heat and humidity freeze conditions based on IEC 61215 [56] standards. Three outcomes are measured to characterize photodegradation: (a) yellowing under UV irradiance, (b) hydrolysis (i.e., moisture induced hazing) and (c) embrittlement. In addition, one sample is retained from further exposure at each time epoch for future investigation.

Fig. 5 summarizes results as yellowing is predominant with exposures consisting of UVA light irradiance while hazing is caused by high level of moisture content and is more predominant when moisture is coupled with UVA light irradiance. Nanoindentation shows that as haze development increases, PET embrittlement results in loss of mechanical properties and physical integrity of the material, which reduces dielectric strength.

4.5. Laboratory: metallization corrosion of bare PV cells

The methodology of degradation science extends beyond bulk material as was applied to accelerated aging of polycrystalline silicon (pc-Si) solar cells, comprising a wealth of functional interfaces of disparate materials with a quantifiable electrical performance rating. The cells, which are a constituent of commercial PV modules, were exposed to weathering conditions in accordance with ASTM G154 [58]. The observed response was maximum power reduction as evidenced by the *I–V* characteristics. These data were mapped onto a macroscopic *physical* model of the solar cell device using an equivalent circuit depiction containing a single diode to represent the photoactive pn junction and parallel and series resistors to represent losses [123]. This simplistic model mathematically predicts the *I–V* characteristics well, and allows for parametrization of the curve for enhanced information extraction and mesoscopic insights to performance loss [124]. In this way the results indicated an increase in the solar cell series resistance that suggests degradation is caused by increased losses at the semiconductor–metal junction. Fig. 6 depicts a series of representative *I–V* curves for a solar cell throughout its exposure. The inset clearly shows that for the sample ensemble the extracted series resistance is increasing in time, which is suggestive that contact resistance of the screen-printed silver conductor lines on the cell are the primary cause of performance loss.

Optical analysis of the screen-printed silver conductor lines that constitute the front contact of the cell indicated corrosion-related oxide nanoparticle formation as evidenced by a decay in the optical second harmonic generation signal [125]. In this way the underly-

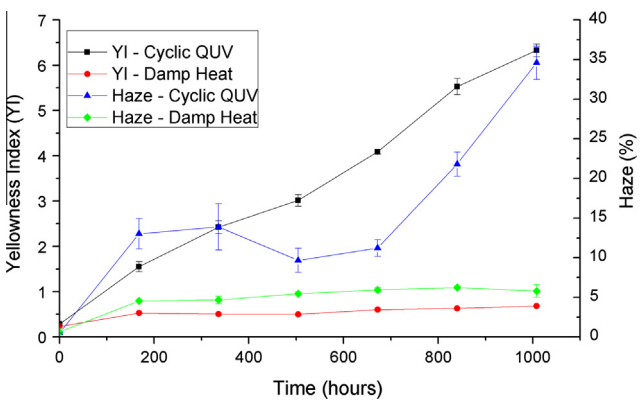

Fig. 5. Degradation of polyethylene-terephthalate (PET): Change in Yellowness Index and Haze (%) with exposure type.

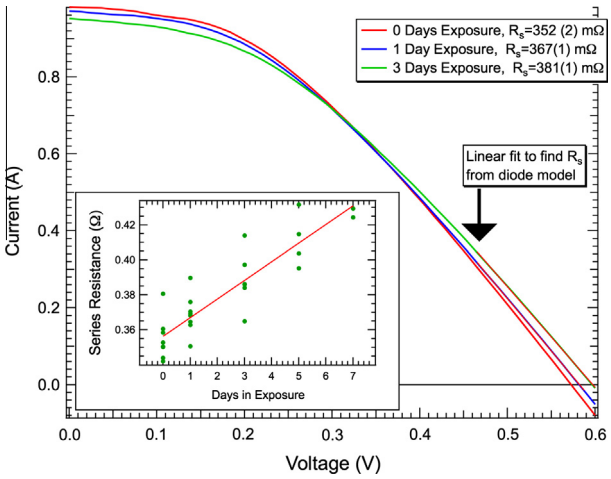

Fig. 6. Acquired current vs voltage characteristics for sample solar cells at various stages of aging exposure. Mapping onto the single diode model allows extraction of a measure of the series resistance, and a plot of series resistance through time (inset) shows an increase commensurate with power loss.

ing *physical* phenomenon of series resistance increase can be probed and understood as part of a mesoscopic model linking the macroscopic diode model. This modeling methodology is useful for confirmatory science and can act as a validation of or constraint upon a semi-gSEM network modeling approach.

4.6. Laboratory: thermal and hydrolytic degradation of PV modules

An example of the semi-gSEM methodology applied to a combination of disparate PV energy materials that comprise a commercial PV module were modeled to find degradation pathways under laboratory-based accelerated aging [89]. The multivariate interactions of mechanistic degradation is complex and best suited for analysis consisting of a network of submodels that give predictive quality. Initial results from full-sized pc-Si PV modules indicated that UV stressors applied according to IEC 61215 UV preconditioning were not sufficient to induce a significant response in the measurable characteristics of the modules within 3000 h. Modules exposed to damp heat from IEC 61215 demonstrated significant degradation within 1890 h, with two dominant mechanistic degradation pathways evident involving several component PV materials. Semi-gSEM analytics of the experimental data, as represented schematically in Fig. 7, indicated that moisture and thermal stressors activate PET hydrolysis (M5) in the pc-Si PV backsheet, and ethylene–vinyl acetate (EVA) encapsulant hydrolysis (M4) in the interior of the module [89]. The rapid equilibration of external environmental moisture levels at these conditions to the interior of the module causes these two degradation pathways to occur in parallel, and appear to be correlated. PET hydrolysis (M5) results in a loss of mechanical, moisture barrier, and dielectric properties in the backsheet as described in Section 4.3. EVA hydrolysis (M4) results in the generation of acetic acid C_{HAc} [126] within the interior of the pc-Si PV module and is strongly correlated to loss of electrical properties (R8 and R9). Integrating the above result on non-encapsulated solar cells the screen-printed silver conductor lines are the suspected root cause of power loss, resulting from the additional acetic acid stressor. (See Fig. 8).

4.7. Real-world: PV module performance study

In order to cross-correlate model prediction to the real world, the performance of a large plurality of PV modules is being acquired for statistical modeling. In an effort to assess and quantify effects of uncontrolled real-world stressors on module performance, potential covariates together with performance measures are collected in a minute-by-minute data stream on 60 pc-Si modules from 20 distinct manufacturers. The stressors and covariates include irradiance levels, ambient temperature, physical module locations and weather conditions such as fog level, cloudiness, snow or rain. The Sunfarm is also continuously monitored to capture unforeseen operation conditions of the PV modules to help capture outliers and to ensure data homogeneity for follow-up studies. In the modeling phase, hierarchical and k-means clustering methods were used to discover naturally arising self similarity groups and performance relationships among the data over time. The data suggested a *statistical* multiple regression model for predicting the real-world module performance in terms of logit transformed power outputs with significant stressors. The findings are again mapped onto physical models in a similar way to the macroscopic diode model described above in Section 4.4. Changes in variances and divergence in performance across a homogeneous study population in time can be used to highlight the nature of damage accumulation, degradation and failure. The population members who fail first can be observed to diverge from the mean population behavior earlier in the study, for a truly non-stochastic degradation pathway.

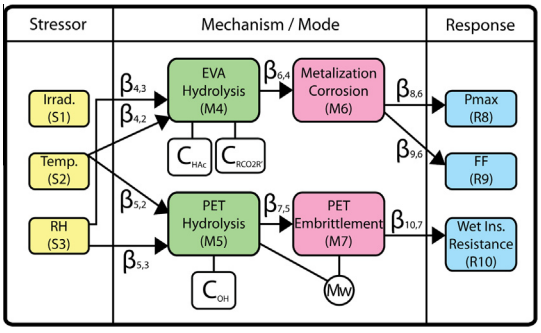

Fig. 7. Two dominant parallel degradation pathways activated during laboratory-based damp heat exposure of PV modules. One pathway is temperature (S2) and humidity (S3) causing hydrolysis of the EVA (M4) encapsulant resulting in acetic acid formation and also temperature (S2) and humidity (S3) causing hydrolysis of the PET (M5) film in the module backsheet. These degradation pathways give rise to both substantial power loss (R8 and R9) and failure of electrical insulation of the backsheet (R10).

4.8. Real-world: microinverter temperature prediction

Ultimately a purely data-driven statistical model can be used to inform the laboratory based studies. A microinverter is typically connected to one PV module to convert the DC output of the PV module to utility AC (Fig. 9(a)). In a similar way to the PV modules, reported data on the microinverter performance naturally allows for extended power electronics informatics and analytics. Fig. 9(c) shows the pairs plot and the correlation coefficient between different environmental, application stressors, and the system response. Furthermore, the direct impact of each stressor on thermal performance are also evaluated. The analysis showed that the critical factors is a mixture of application stressors: PV module temperature, AC power output of the microinverter, and environmental stressors: irradiance and ambient temperature [127]. A multiple linear regression model was developed to predict the noon time microinverter temperature at real-world operation. Fig. 9(b) shows comparison between actual and predicted microinverter temperature during noon time in a typical cloudy day.

Even in this complex device, a switch-mode power converter, the microstructural evolution of the power components results in performance degradation at the mesoscale, but the stressors and responses are largely unknown. Application stressors, such as high speed grid disturbances, high frequency switching power cycling, combined with slower events such as module power cycling and heat/humidity give a rich set of variables that impact inverter mesostructural evolution. By acquiring time-series datasets, typically on the minute-by-minute level, one can identify factors through which some population members diverge from the mean behavior and become a subset of extremal or even failing devices. One seeks out the relationships between these variables and again, mapping the data onto *physical* models allows close examination of what is essentially a closed device. For example in MOSFETs utilized in microinverters *physical* models of efficiency loss indicate that switching losses have two contributors: output capacitance loss, $P_{oss} = \frac{1}{2} C_{oss} V^2 f_s$ and conduction losses $P_{cl} = I_D^2 R_{on}$, regression fitting determines the *physical* parameters of the device and allow for monitoring in time to infer degradation mechanisms, such as an R_{on} increase signifying thermal runaway of the FET die. Mechanisms can then be confirmed by laboratory-based analytic investigation. In many ways, this is easier and more productive way to determine the critical factors on mesoscopic evolution over lifetime. Difference can arise from climatic, application stressors, design, construction, etc.

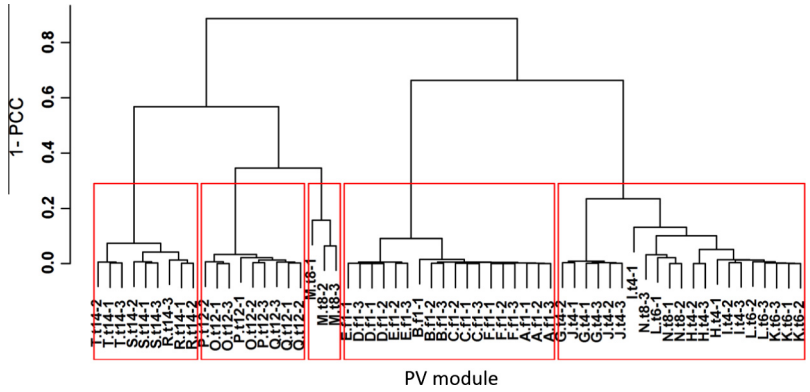

Fig. 8. The hierarchical clustering dendrogram of solar noon time Performance Ratio (PR) [57] time series of 60 pc-Si modules. Each sample is labeled as "Brands (A - T).Location (tracker or fixed) – Sample (1–3)". The height of the dendrogram is computed as "1 – PCC (Pearsons Correlation Coefficient). PCC indicates the similarity of two samples or clusters. The number of clusters is chosen through elbow-point method by K-means clustering.

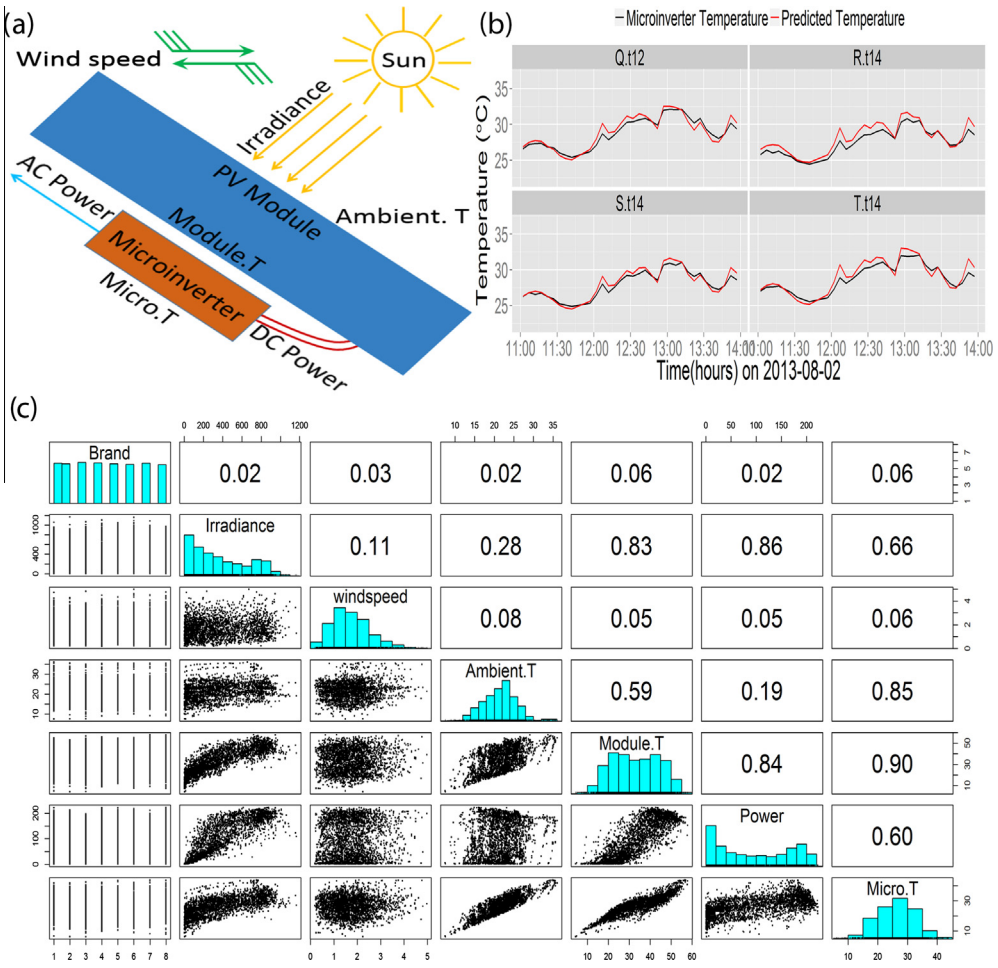

Fig. 9. (a) Schematic diagram of PV module and microinverter setup. (b) Comparison of actual microinverter temperature and fitted microinverter temperature for the microinverters connected to four different PV module brands during noon time on a typical cloudy day. (c) Pairs plot and correlation coefficient between different environmental and application stressors. Irradiance, wind speed and ambient temperature (Ambient.T) are the environmental stressors. PV module temperature (Module.T), PV module brand(Brand), AC power (Power) and microinverter temperature (Micro.T) are application stressors.

5. Challenges and paths forward for degradation science

Degradation science as outlined in this paper poses several new challenges, complementing traditional research approaches with new epidemiological, analytical, modeling and visualization algorithms, including harnessing diverse data provenience and formats as well as de novo data types for research, with experiments conducted under broad and uncontrolled conditions. The real-world massive data input and processing naturally enforces sharing of research data [128] and processing codes [44,45], as well as promotes open data, open science, and open government policies [41,42].

5.1. The need for a shared energy materials ontology

Standard methods developed for single scale, separate, or moderate-sized datasets do not scale to modern big data from multiple sources. Six common challenges are recognized by the National Research Councils Committee on the Analysis of Massive Data [38]. (a) Tracking data provenance. The context in which the original data has been collected needs to be carefully formulated, managed, and shared with original data. Data about data, or metadata, is a crucial aspect of data annotation and curation that will affect the aggregation and sharing of data from multiple heterogeneous sources. (b) Coping with sampling biases and heterogeneity. Energy material data may come from (hidden) heterogeneous groups, and contain outliers/corruptions, measurement errors, missing values or be subject to selection biases. It is important to develop methods to extract true signals from noisy and incomplete data. (c) Working with different data formats and structures. Because of the variety of energy material data originating from different vendors an important aspect of informatics work would be to map and align such data to ensure that semantics is preserved during data transmission, transformation and integration. (d) Scalability of computational algorithms for processing and analyzing energy materials data. Advances in cloud-computing have provided a promising computational paradigm for scale up computational speed. However, specific MapReduce algorithms for specific application tasks are yet to be developed. (e) Ensuring data integrity, data security, and enabling data integration and data sharing. This has been incorporated in the design of CRADLE™ and will be fully developed and deployed. (f) Data visualization. Methods for visualizing massive energy materials data need to be further developed for on-the-fly analytics and degradation pathway network decision-making.

To address these challenges the energy materials community needs to develop a specific ontology that captures the scope and depth of the energy materials domain. An energy materials ontology (EMO) can serve as a foundation for addressing the challenges of data provenance, data variety, and facilitate data integration and sharing. Such an approach has been proven imperative for the biological research community as represented by the Gene Ontology [129] for disease specific consortia such as epilepsy research community [130] or the ontological resources provided by the National Library of Medicine encapsulated in the Unified Medical Language System [131].

The important and direct roles a domain ontology can play for managing big data has been demonstrated in [132], where a common ontology directly supports data capture, data integration, data sharing, and user interfaces for data retrieval.

5.2. Big data opportunities in PV energy materials

Big data is commonly characterized by variety, volume, velocity and veracity [133]. Progress in distinct domains may require different strategies. It is important to resist the temptation of a monolithic approach [134].

Two recent national center projects from the domain of biomedicine offer a blueprint for degradation science in its approach to creating a community-level data and informatics infrastructure. The first is the National Sleep Research Resource (NSRR [135]), which offers free web access to large retrospective collections of de-identified physiological signals and clinical data elements collected in well-characterized research cohorts and clinical trials. The underlying architecture is motivated from the PhysioMIMI project, in its ontology-driven architecture governing data curation, data integration, data access and data analytics. The second is Center for SUDEP Research (CSR [136]), which adopts MEDCIS as its data and informatics platform, where the purpose is to prospectively follow a larger group of epilepsy patients at an elevated risk for SUDEP overtime, from multiple hospital Epilepsy Monitoring Units. Again, a dedicated Epilepsy and Seizure Ontology serves as the core knowledge source driving the entire data lifecycle. The general ontology-guided approach to developing information systems is discussed in a larger context in [137]. In both of the exemplar projects, big data challenges such as processing and annotating large collection of physiological signals and managing patient records with thousands of uneven attribute columns unfolded themselves from the agile development process and computation need, motivated by the projects grander vision.

In addition to an ontology-driven system architecture, context-aware, role-based access management, user-centered, interface-driven development and agile process in partnership with domain experts are key elements for progress. A true collaborative partnership between computer scientists and domain experts, in the spirit of team science, sets a solid organizational foundation for projects such as NSRR and CSR.

In the context of degradation studies, we refer to *"Big Data as a frame of mind, or a bigger vision,* in perceiving the science and engineering landscape from a grander data scale, emboldened by the scalability of cloud computing, such as MapReduce for massive parallel processing" [138]. Such an approach can dramatically accelerate the speed of analysis in cases of complex tasks that are previously less computationally feasible [139]. We believe that such a scalable approach is beneficial for degradation research in general, even for computationally feasible problems, because it allows us to ask bigger questions and to answer them faster, putting the computational barrier in the back of our minds so we can focus more on the scientific content.

5.3. Analytics for hypothesis-driven and epidemiological research

A challenge to development of a holistic network of submodels fully characterizing the time-evolution of material is the integration of models and datasets from disparate experiments and the merging of hypothesis-driven research and epidemiological research. Examples of laboratory based studies in a stress/mechanism-mode/response framework are often initially guided by hypothesis, as in a semi-supervised fashion, and then a data-driven mode mesoscopic model is constructed and informed by existing physical models. Real-world studies by contrast are more clearly epidemiological in structure and lend themselves to predictive modeling based purely in statistics. However, as we've shown, the projection of these statistical models onto physical models can provide fundamental insights into mechanisms of degradation and serve to validate and verify the predictions given by laboratory-based studies or act as regression constraints.

Relevant statistical analytics need not only overcome the heterogeneous nature of different types/evolution of data, but also the high/huge dimensions resultants from multiple factors and measurements obtained on the same module/material over time, as well as individual patterns that can only be learned more accurately as data are updated under each particular environment. The heterogeneous real-life environment can lead to measurement errors, irregular sampling points and censoring (as we have seen in our SunFarm study [140]). Thus analytics that lead to effective measures that are fast to compute, effective for prediction and well-conditioned under this heterogeneity will be important as they are to classification of huge volumes of astronomical light curves [141]. These measures are also analogous to descriptors used in image analyses. Individualized longitudinal predictive models that can be updated over time [142] will be extremely important for critical system monitoring. The software packages for effective change point analyses that are applicable to degradation science and monitor system transitions are needed [143,144].

5.4. Encompassing and discriminating temporal domains

The semi-gSEM results to date contain no explicit time dependency in the mesoscopic evolution, although this time dependency is an eventual necessity for a complete picture of material degradation. A time dependency in the modeling must discriminate the dynamics of typical function, and be sensitive to the material mesoscopic evolution related to degradation. The temporal analytics must have the sensitivity to distinguish and discover the potential slow and rare events associated with degradation and failure. These concepts require high density datastreams as inputs to the model so that information entropy is not limiting to the discovery science. The segmentation of time-scales into cycles (daily variation), steps (functional degradation response from stressors), and stages (phase transformations) enables temporal discrimination. Temporal discrimination among dynamics, function and the slow or rare degradation mechanisms and modes is achieved through data filtering, dimensional reduction or frequency banding using spectral time-series methods [145], with the caveat that data are always retained so that the filtering can be applied repetitively as more data become available. This temporal segmentation is similar conceptually to the analysis of Brownian motion by Einstein, where the physical picture of Brownian motion is described as a function of collective, stochastic phenomena governing the ensemble of particles to the motion of a single particle on a larger order of time, given by the square root of elapsed time [146].

The inclusion of time is a specific challenge due to the bridging of multiple temporal orders of magnitude from the initiation and accumulation of damage through the phase transformation leading to cascading macroscopic performance loss on a decade time-scale. These timescales of transformation are mechanistically dependent, as depicted in Fig. 1, and the network of submodels must be inclusive of these various time stages. Phase transformations are not modeled by a priori physical models to an exact degree, and it is probable that the supervised or statistical validation provided by physical models based in hypothesis-driven science are valid within these time stages, and they must be linked by parametric statistical models such as regression splines. The regression splines can effectively model these temporal change points and ensure a continuity of the predictive model.

5.5. Making the stochastic deterministic

Integrating laboratory-based and real-world data-driven studies by necessity implies not only bridging micro-, meso-, and macro-scopic spatio-temporal scales, but also connecting the deterministic and stochastic approaches to modeling [147]. Stochastic reaction–diffusion systems [148] are an example of this type of connection in systems that, while being described by deterministic equations, also allow for intrinsic fluctuations that turn out to be an important factor leading to drastic changes in the dynamics [149]. Degradation mesocale evolution, allowing for externally induced structural disorder of either annealed or quenched type, would introduce an additional distinct level of coupling between thermal fluctuations and external stressor generated disorder that has never been explored before.

The interplay of stochastic and deterministic evolution, informed by semi-gSEM merging of *physical* and *statistical* models, should help us in identifying the deterministic part of the mesoscale evolution dynamics on one hand, but also the effects of structural disorder stemming from the externally imposed stressor fluctuations on the other. This is of course a challenge that any modeling of the degradation phenomena will have to face in one way or another. To us data-driven *statistical* modeling coupled to multi-level multi-scale *physical* models, bridging across gaps of approximations, is the way to address this challenge.

6. Conclusion

Based on recent advances in nanoscience, data science and the availability of massive real-world datastreams, the mesoscopic evolution of energy materials can now be more fully studied. The temporal evolution is vastly complex in time and length scales and is fundamentally challenging to scientific understanding of degradation mechanisms and pathways responsible for energy materials evolution over lifetime. We propose a paradigm shift towards mesoscopic evolution modeling, based on physical and statistical models, that would integrate laboratory studies and real-world massive datastreams into a stress/mechanism/response framework with predictive capabilities. These epidemiological studies encompass the variability in properties that affect performance of material ensembles. Mesoscopic evolution modeling is shown to encompass the heterogeneity of these materials and systems, and enables the discrimination of the fast dynamics of their functional use and the slow and/or rare events of their degradation. We delineate paths forward for degradation science.

Acknowledgments

The authors would like to acknowledge the funding for this work. The SDLE center was established through funding through the Ohio Third Frontier, Wright Project Program Award Tech 12-004. The acrylic research was supported by Ohio Third Frontier Award Tech 11-060 in collaboration with Replex Plastics. The PV module damp heat study was supported by Underwriters Laboratories. The real-world PV module studies are supported the Bay Area Photovoltaic Consortium Prime Award No. DE-EE0004946, Subaward Agreement No. 60220829-51077-T. Sample tracking was provided by REDCap [150] funded by the Clinical and Translational Science Collaborative (CTSC) under Grant No. UL1TR000439. Additional funding and support from DuPont, Saint Gobain, 3M, SunPower, Solexel, Kent Displays, MCCO Inc., Q-Lab, Underwriters Laboratories and IIT-GN is acknowledged. Additionally, the authors acknowledge the thoughtful and helpful comments of the reviewers.

References

[1] Williard N, He Wei, Hendricks C, Pecht M. Lessons learned from the 787 Dreamliner issue on lithium-ion battery reliability. Energies (19961073) 2013;6(9):4682–95. http://dx.doi.org/10.3390/en6094682. <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=90506546&site=ehost-live>.

[2] Huerta MPH. Certification testing of lithium-ion batteries to be used on commercial airplanes – ntsb safety reccomendation, Tech. Rep. A-14-032 through -036, May 2014 <http://www.ntsb.gov/doclib/recletters/2014/A-14-032-036.pdf>.

[3] P. Knudson, NTSB recommends process improvements for certifying lithium-ion batteries as it concludes its investigation of the 787 boston battery fire incident; 2014 <http://www.ntsb.gov/news/2014/141201.html>.

[4] Ross RG. PV reliability development lessons from JPL's flat plate solar array project. IEEE J Photovolt 2014;4(1):291–8. http://dx.doi.org/10.1109/JPHOTOV.2013.2281102. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6616584>.

[5] Köntges M, Kurtz S, Packard C, Jahn U, Berger K, Kato K, et al. IEA-PVPS task 13: review of failures of PV modules. Tech. rep., May 2014 <http://iea-pvps.org/index.php?id=275>.

[6] Poelking C, Tietze M, Elschner C, Olthof S, Hertel D, Baumeier B, et al. Impact of mesoscale order on open-circuit voltage in organic solar cells. Nature Mater, http://dx.doi.org/10.1038/nmat4167 <http://www.nature.com/nmat/journal/vaop/ncurrent/full/nmat4167.html>.

[7] Hemminger J, Crabtree G, Sarrao J. From quanta to the continuum: Opportunities for mesoscale science. Tech. rep., US Department of Energy Basic Energy Sciences Advisory Committee, September 2012.

[8] Hemminger J, Crabtree G, Malozemoff A. Science for energy technology: strengthening the link between basic research and industry. Tech. rep.; 2010 <http://science.energy.gov/bes/news-and-resources/reports/abstracts/#SET>.

[9] Chen RS. A vast machine: computer models, climate data, and the politics of global warming. Environ Health Perspect 2011;119(4):A182. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3080963/>.

[10] Edwards PN. A vast machine: computer models, climate data, and the politics of global warming. 1st ed. Cambridge (Mass): The MIT Press; 2010.

[11] Roco MC. The long view of nanotechnology development: the national nanotechnology initiative at 10 years. In: Nanotechnology research directions for societal needs in 2020, science policy reports. Netherlands: Springer; 2011. p. 1–28. <http://link.springer.com/chapter/10.1007/978-94-007-1168-6_1>.

[12] Abbamonte P. Condensed-matter physics: picking up fine vibrations. Nature Phys 2013;9(12):759–60. http://dx.doi.org/10.1038/nphys2832. <http://www.nature.com/nphys/journal/v9/n12/full/nphys2832.html>.

[13] Hickstein DD, Dollar F, Ellis JL, Schnitzenbaumer KJ, Keister KE, Petrov GM, et al. Mapping nanoscale absorption of femtosecond laser pulses using plasma explosion imaging. ACS Nano 2014;8(9):8810–8. http://dx.doi.org/10.1021/nn503199v. <http://pubs.acs.org/doi/abs/10.1021/nn503199v>.

[14] Binder K, Kob W. Glassy materials and disordered solids: an introduction to their statistical mechanics. World Scientific; 2011.

[15] Unold T, Schock H. Nonconventional (non-silicon-based) photovoltaic materials. Ann Rev Mater Res 2011;41(1):297–321. http://dx.doi.org/10.1146/annurev-matsci-062910-100437. 00028 <http://dx.doi.org/10.1146/annurev-matsci-062910-100437>.

[16] Dai L, Chang DW, Baek J-B, Lu W. Carbon nanomaterials for advanced energy conversion and storage. Small 2012;8(8):1130–66. http://dx.doi.org/10.1002/smll.201101594. 0022 <http://onlinelibrary.wiley.com/doi/10.1002/smll.201101594/abstract>.

[17] Verma P, Maire P, Novák P. A review of the features and analyses of the solid electrolyte interphase in Li-ion batteries. Electrochim Acta 2010;55(22):6332–41. http://dx.doi.org/10.1016/j.electacta.2010.05.072. <http://linkinghub.elsevier.com/retrieve/pii/S0013468610007747>.

[18] Koumoto K, Wang Y, Zhang R, Kosuga A, Funahashi R. Oxide thermoelectric materials: a nanostructuring approach. Ann Rev Mater Res 2010;40(1):363–94. http://dx.doi.org/10.1146/annurev-matsci-070909-104521. <http://dx.doi.org/10.1146/annurev-matsci-070909-104521>.

[19] French R, Parsegian V, Podgornik R, Rajter R, Jagota A, Luo J, et al. Long range interactions in nanoscale science. Rev Mod Phys 2010;82(2):1887–944. http://dx.doi.org/10.1103/RevModPhys.82.1887. <http://apps.isiknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=1&SID=1Ah3Ejhdn8kKF83ia@F&page=1&doc=2>.

[20] French R, Tran H. Immersion lithography: photomask and wafer-level materials. Ann Rev Mater Res 2009;39:93–126. http://dx.doi.org/10.1146/annurev-matsci-082908-145350. <http://apps.isiknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=1&SID=1Ah3Ejhdn8kKF83ia@F&page=1&doc=7>.

[21] Karmakar S, Dasgupta C, Sastry S. Growing length scales and their relation to timescales in glass-forming liquids. Ann Rev Cond Matter Phys 2014;5(1):255–84. http://dx.doi.org/10.1146/annurev-conmatphys-031113-133848. <http://dx.doi.org/10.1146/annurev-conmatphys-031113-133848>.

[22] Duoss E, Zhu C, Sullivan K, Vericella J, Hopkins J, Zheng R, et al. Additive micro-manufacturing of designer materials. In: Udomkichdecha W, Bllinghaus T, Manonukul A, Lexow J, editors. Materials challenges and testing for manufacturing, mobility, biomedical applications and climate. Springer International Publishing; 2014. p. 13–24. 00000 <http://link.springer.com/chapter/10.1007/978-3-319-11340-1_2>.

[23] Sun K, Wei T-S, Ahn BY, Seo JY, Dillon SJ, Lewis JA. 3D printing of interdigitated Li-ion microbattery architectures. Adv Mater 2013;25(33):4539–43. http://dx.doi.org/10.1002/adma.201301036. 0004 http://doi.wiley.com/10.1002/adma.201301036.

[24] Hartgerink JD, Beniash E, Stupp SI. Self-assembly and mineralization of peptide-amphiphile nanofibers. Science 2001;294(5547):1684–8. http://dx.doi.org/10.1126/science.1063187. <http://www.sciencemag.org/content/294/5547/1684>.

[25] Sewell H, Mulkens J. Materials for optical lithography tool application. Ann Rev Mater Res 2009;39(1):127–53. http://dx.doi.org/10.1146/annurev-matsci-082908-145309. <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=4AHB4MbNNLdDl939HC6&page=2&doc=15>.

[26] Macak JM, Tsuchiya H, Ghicov A, Yasuda K, Hahn R, Bauer S, et al. TiO_2 nanotubes: self-organized electrochemical formation, properties and applications. Curr Opin Solid State Mater Sci 2007;11(12):3–18. http://dx.doi.org/10.1016/j.cossms.2007.08.004. <http://www.sciencedirect.com/science/article/pii/S1359028607000496>.

[27] Yang L, Adam C, Nichol GS, Cockroft SL. How much do van der waals dispersion forces contribute to molecular recognition in solution? Nature Chem 2013;5(12):1006–10. http://dx.doi.org/10.1038/nchem.1779. <http://www.nature.com/doifinder/10.1038/nchem.1779>.

[28] Garnett EC, Brongersma ML, Cui Y, McGehee MD. Nanowire solar cells. Ann Rev Mater Res 2011;41(1):269–95. http://dx.doi.org/10.1146/annurev-matsci-062910-100434. <http://dx.doi.org/10.1146/annurev-matsci-062910-100434>.

[29] Crabtree G, Glotzer SC, McCurdy B, Roberto J. Computational materials science and chemistry: accelerating discovery and innovation through simulation-based engineering and science; 2010 <http://science.energy.gov/bes/news-and-resources/reports/abstracts/#CMSC>.

[30] Brenner DW. Challenges to marrying atomic and continuum modeling of materials. Curr Opin Solid State Mater Sci 2013;17(6):257–62. http://dx.doi.org/10.1016/j.cossms.2013.07.005. <http://www.sciencedirect.com/science/article/pii/S1359028613000533>.

[31] Praprotnik M, Site LD, Kremer K. Multiscale simulation of soft matter: from scale bridging to adaptive resolution. Ann Rev Phys Chem 2008;59(1):545–71. http://dx.doi.org/10.1146/annurev.physchem.59.032607.093707. http://dx.doi.org/10.1146/annurev.physchem.59.032607.093707.

[32] Kamerlin SC, Vicatos S, Dryga A, Warshel A. Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems. Ann Rev Phys Chem 2011;62(1):41–64. http://dx.doi.org/10.1146/annurev-physchem-032210-103335. http://dx.doi.org/10.1146/annurev-physchem-032210-103335.

[33] Kevrekidis IG, Samaey G. Equation-free multiscale computation: algorithms and applications. Ann Rev Phys Chem 2009;60(1):321–44. http://dx.doi.org/10.1146/annurev.physchem.59.032607.093610. http://dx.doi.org/10.1146/annurev.physchem.59.032607.093610.

[34] Walpole J, Papin JA, Peirce SM. Multiscale computational models of complex biological systems. Ann Rev Biomed Eng 2013;15(1):137–54. http://dx.doi.org/10.1146/annurev-bioeng-071811-150104. http://dx.doi.org/10.1146/annurev-bioeng-071811-150104.

[35] Holdren J. Goals of the materials genome initiative j the white house; 2011 <http://www.whitehouse.gov/mgi/goals>.

[36] de Pablo JJ, Jones B, Kovacs CL, Ozolins V, Ramirez AP. The materials genome initiative, the interplay of experiment, theory and computation. Curr Opin Solid State Mater Sci 2014;18(2):99–117. http://dx.doi.org/10.1016/j.cossms.2014.02.003. <http://www.sciencedirect.com/science/article/pii/S1359028614000060>.

[37] Andriotis AN, Mpourmpakis G, Broderick S, Rajan K, Datta S, Sunkara M, et al. Informatics guided discovery of surface structure–chemistry relationships in catalytic nanoparticles. J Chem Phys 2014;140(9):094705. http://dx.doi.org/10.1063/1.4867010. <http://scitation.aip.org/content/aip/journal/jcp/140/9/10.1063/1.4867010>.

[38] in Jordan M, editor. Frontiers in massive data analysis. National research council. National Academies Press; 2013. <http://www.nap.edu/catalog.php?record_id=18374>.

[39] Obama B. US OSTP executive order – making open and machine readable the new default for government information j the White House; 2013 <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->.

[40] Charter OD. Open data charter <https://www.gov.uk/government/publications/open-data-charter>, June 2013 <https://www.gov.uk/government/publications/open-data-charter>.

[41] Holdren JP. Promoting open data, open science, and open government; 2014 <http://www.whitehouse.gov/administration/eop/ostp/initiatives#Openness>.

[42] Hendler J, Holm J, Musialek C, Thomas G. US government linked open data: Semantic.data.gov. IEEE Intell Syst 2012;27(3):25–31.

[43] Maxwell E. Harnessing openness to improve research, teaching and learning in higher education. Innov: Technol, Govern, Global 2010;5(2):155–66. http://dx.doi.org/10.1162/inov_a_00019. http://dx.doi.org/10.1162/inov_a_00019.

[44] Ince DC, Hatton L, Graham-Cumming J. The case for open computer programs. Nature 2012;482(7386):485–8. http://dx.doi.org/10.1038/nature10836. <http://www.nature.com/nature/journal/v482/n7386/full/nature10836.html>.

[45] Plimpton SJ, Gale JD. Developing community codes for materials modeling. Curr Opin Solid State Mater Sci 2013;17(6):271–6. http://dx.doi.org/10.1016/j.cossms.2013.09.005. <http://www.sciencedirect.com/science/article/pii/S1359028613000764>.

[46] Ghemawat S, Gobioff H, Leung S-T. The Google file system. ACM SIGOPS Operat Syst Rev 2003;37(5):29–43. http://dx.doi.org/10.1145/1165389.945450. <http://journals.ohiolink.edu/ejc/article.cgi?issn=01635980&issue=v37i0005&article=29_tgfs>.

[47] Chang F, Dean J, Ghemawat S, Hsieh WC, Wallach DA, Burrows M, et al. Bigtable: a distributed storage system for structured data. ACM Trans Comp Syst (TOCS) 2008;26(2):1–26. http://dx.doi.org/10.1145/1365815.1365816. <http://journals.ohiolink.edu/ejc/article.cgi?issn=07342071&issue=v26i0002&article=1_b>.

[48] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Commun ACM 2008;51(1):107–13. http://dx.doi.org/10.1145/1327452.1327492. 11882, <http://journals.ohiolink.edu/ejc/article.cgi?issn=00010782&issue=v51i0001&article=107_m>.

[49] de Graef M. Storing, archiving, and retrieving materials data sets. Ann Rev Mater Res 2015;45(1). http://dx.doi.org/10.1146/annurev-matsci-070214-020844. null, 00000 <http://www.annualreviews.org/doi/abs/10.1146/annurev-matsci-070214-020844>.

[50] Goodenough JB, Kim Y. Challenges for rechargeable li batteries. Chem Mater 2010;22(3):587–603. http://dx.doi.org/10.1021/cm901452z. http://dx.doi.org/10.1021/cm901452z.

[51] Hacke P, Terwilliger K, Smith R, Glick S, Pankow J, Kempe M, et al. System voltage potential-induced degradation mechanisms in PV modules and methods for test. In: 37th IEEE photovoltaic specialists conference (PVSC). IEEE; 2011. p. 000814–20. <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6186079>.

[52] Novoa FD, Miller DC, Dauskardt RH. Environmental mechanisms of debonding in photovoltaic backsheets. Solar Energy Mater Solar Cells 120, Part A 2014:87–93. http://dx.doi.org/10.1016/j.solmat.2013.08.020. <http://www.sciencedirect.com/science/article/pii/S0927024813004194>.

[53] Peike C, Hoffmann S, Hülsmann P, Weiß K-A, Koehl M, Bentz P. Permeation impact on metallization degradation 2012;84720U-1–84720U-9, http://dx.doi.org/10.1117/12.929819 <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1379672>.

[54] Koehl M, Heck M, Wiesmeier S. Modelling of conditions for accelerated lifetime testing of humidity impact on PV-modules based on monitoring of climatic data. Solar Energy Mater Solar Cells 2012;99:282–91. http://dx.doi.org/10.1016/j.solmat.2011.12.011. <http://journals.ohiolink.edu/ejc/article.cgi?issn=09270248&issue=v99inone_c&article=282_mocfalbomocd>.

[55] Kempe MD, Panchagade D, Reese MO, Dameron AA. Modeling moisture ingress through polyisobutylene-based edge-seals. Progr Photovolt: Res Appl 2014;n/a–n/a, http://dx.doi.org/10.1002/pip.2465 <http://onlinelibrary.wiley.com/doi/10.1002/pip.2465/abstract>.

[56] IEC, IEC61215 ed2.0 – Crystalline silicon terrestrial photovoltaic (PV) modules – design qualification and type approval j IEC webstore j publication abstract, preview, scope; 2005 <http://webstore.iec.ch/webstore/webstore.nsf/Artnum_PK/34077>.

[57] IEC 61724 ed1.0 - Photovoltaic system performance monitoring – guidelines for measurement, data exchange and analysis j IEC webstore j publication abstract, preview, scope; 1998 <http://webstore.iec.ch/webstore/webstore.nsf/Artnum_PK/22954>.

[58] G03 Committee. Practice for operating fluorescent light apparatus for UV exposure of nonmetallic materials. Tech. Rep. G154-06, ASTM International; 2006 <http://www.astm.org/doiLink.cgi?G154>.

[59] Osterwald CR, McMahon TJ. History of accelerated and qualification testing of terrestrial photovoltaic modules: a literature review. Progr Photovolt: Res Appl 2009;17(1):11–33. http://dx.doi.org/10.1002/pip.861. http://doi.wiley.com/10.1002/pip.861.

[60] Hasselbrink E, Anderson M, Defreitas Z, Mikofski M, Shen Y-C, Caldwell S, et al. Validation of the PVLife model using 3 million module-years of live site data. In: IEEE 39th photovoltaic specialists conference (PVSC). IEEE; 2013. p. 0007–12. <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6744087>.

[61] Stein JS, King BH. Modelling for PV plant optimization. Photovolt Int 19th Ed – J Solar Media Store (PVI) 2013;19:101–9. 00000 <http://www.solarmediastore.com/journals/pvi-19.html>.

[62] Hu Y. PV module performance under real-world test conditions – a data analytics approach, Master of science thesis, Case Western Reserve University; 2014 <https://etd.ohiolink.edu/ap/10?0::NO:10:P10_ACCESSION_NUM:case1396615109>.

[63] Hossain MA. Thermal characteristics of microinverters on dual-axis trackers, Master of science thesis, Case Western Reserve University, Cleveland, OH; 2014 <http://rave.ohiolink.edu/etdc/view?acc_num=case1396888841>.

[64] Jordan DC, Kurtz SR. Photovoltaic degradation rates-an analytical review: photovoltaic degradation rates. Progr Photovolt: Res Appl 2013;21(1):12–29. http://dx.doi.org/10.1002/pip.1182. http://doi.wiley.com/10.1002/pip.1182.

[65] Janakeeraman SV, Singh J, Kuitche J, Mallineni JK, TamizhMani G. A statistical analysis on the cell parameters responsible for power degradation of fielded PV modules in a hot-dry climate. In: 40th Photovoltaic specialist conference (PVSC), 2014 IEEE. IEEE, Denver, CO; 2014. p. 3234–8, http://dx.doi.org/10.1109/PVSC.2014.6925624 <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6925624>.

[66] Guttorp P. Statistics and climate. Ann Rev Statist Appl 2014;1(1):87–101. http://dx.doi.org/10.1146/annurev-statistics-022513-115648. http://dx.doi.org/10.1146/annurev-statistics-022513-115648.

[67] Randall D, Khairoutdinov M, Arakawa A, Grabowski W. Breaking the cloud parameterization deadlock. Bullet Am Meteor Soc 2003;84(11):1547–64. http://dx.doi.org/10.1175/BAMS-84-11-1547. <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-84-11-1547>.

[68] in Rajan K, editor. Informatics for materials science and engineering. Oxford: Butterworth-Heinemann; 2013. <http://www.sciencedirect.com/science/article/pii/B9780123943996000217>.

[69] Rajan K. Materials informatics. Ann Rev Mater Res 45(1), http://dx.doi.org/10.1146/annurev-matsci-070214-021132 <http://www.annualreviews.org/doi/abs/10.1146/annurev-matsci-070214-021132>.

[70] Gillespie DT. Stochastic simulation of chemical kinetics. Ann Rev Phys Chem 2007;58(1):35–55. http://dx.doi.org/10.1146/annurev.physchem.58.032806.104637. http://dx.doi.org/10.1146/annurev.physchem.58.032806.104637.

[71] Lohr SL. Sampling: design and analysis, 2nd ed. Boston (Mass): Cengage Learning; 2009.

[72] Jackson M, Cox D. The principles of experimental design and their application in sociology. Ann Rev Sociol 2013;39(1):27–49. http://dx.doi.org/10.1146/annurev-soc-071811-145443. http://dx.doi.org/10.1146/annurev-soc-071811-145443.

[73] Pearl J. Causality. Cambridge University Press; 2009. <http://books.google.com/books?id=f4nuexsNVZIC>.

[74] Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. 2nd ed. Hoboken (NJ): Wiley; 2011.

[75] Murray MP, Bruckman LS, French RH. Photodegradation in a stress and response framework: poly(methyl methacrylate) for solar mirrors and lens. J Photon Energy 2012;2(1). http://dx.doi.org/10.1117/1.JPE.2.022004. 022004–022004, http://dx.doi.org/10.1117/1.JPE.2.022004.

[76] Ching W-Y, Rulis P. Electronic structure methods for complex materials: the orthogonalized linear combination of atomic orbitals. Oxford: Oxford University Press; 2012.

[77] Chaikin PM, Lubensky TC. Principles of condensed matter physics. Cambridge (New York, NY, USA): Cambridge University Press; 2000 [reprint ed.].

[78] Gennes PGd, Prost J. The physics of liquid crystals. 2nd ed. Oxford; New York: Oxford University Press; 1995.

[79] Kleman M, Laverntovich OD, Friedel J. Soft matter physics: an introduction. 2003rd ed. New York: Springer; 2002.

[80] Happel J, Brenner H. Low Reynolds number hydrodynamics: with special applications to particulate media. 1983rd ed. The Hague; Boston: Hingham, MA, USA: Springer; 1983.

[81] ben Avraham D, Havlin S. Diffusion and reactions in fractals and disordered systems. Cambridge: Cambridge University Press; 2005.

[82] Cavagna A, Giardina I. Bird flocks as condensed matter. Ann Rev Condensed Matter Phys 2014;5(1):183–207. http://dx.doi.org/10.1146/annurev-conmatphys-031113-133834. http://dx.doi.org/10.1146/annurev-conmatphys-031113-133834.

[83] Garner MG, Hamilton SA. Principles of epidemiological modelling. Rev Scient Tech (Int Off Epizoot) 2011;30(2):407–16. <http://europepmc.org/abstract/med/2196121>.

[84] Nelson DR. Defects and geometry in condensed matter physics. Cambridge (New York): Cambridge University Press; 2002.

[85] Binder K, Kob W. Glassy materials and disordered solids: an introduction to their statistical mechanics. Hackensack (NJ): World Scientific Publishing Company; 2005.

[86] Weinberger D. The machine that would predict the future. Scient Am 2011;305(6):52–57, 0001. http://dx.doi.org/10.1038/scientificamerican1211-52. <http://www.nature.com/scientificamerican/journal/v305/n6/full/scientificamerican1211-52.html>.

[87] Steinhauser MO. Computational multiscale modeling of fluids and solids, vol. 60. Germany: Springer; 2008. p. 00035.

[88] Merchant BA, Madura JD. A review of coarse-grained molecular dynamics techniques to access extended spatial and temporal scales in biomolecular simulations. Ann Rep Comput Chem 2011;7:67–87.

[89] Bruckman LS, Wheeler NR, Ma J, Wang E, Wang CK, Chou I, et al. Statistical and domain analytics applied to PV module lifetime and degradation science. IEEE Access 2013;1:384–403. http://dx.doi.org/10.1109/ACCESS.2013.2267611. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6527980>.

[90] Taleb NN. The black swan: the impact of the highly improbable. New York (NY, USA): Random House Trade Paperbacks; 2010.

[91] Taleb NN. Antifragile: things that gain from disorder. New York (NY, USA): Random House; 2012.

[92] Naji A, Ghodrat M, Komaie-Moghaddam H, Podgornik R. Asymmetric coulomb fluids at randomly charged dielectric interfaces: anti-fragility. Overcharg Charge Invers 2014;141:174704. preprint arXiv:1409.2609 <http://arxiv.org/abs/1409.2609>.

[93] Ramirez ALB, Kean ZS, Orlicki JA, Champhekar M, Elsakr SM, Krause WE, et al. Mechanochemical strengthening of a synthetic polymer in response to typically destructive shear forces. Nature Chem 2013;5(9):757–61. http://dx.doi.org/10.1038/nchem.1720. <http://www.nature.com/doifinder/10.1038/nchem.1720>.

[94] Zhang G-Q, Mueller RSW, Szymanski J, Troy A, Wilson DL, Flask CA, et al. Multi-modality, multi-resource, information integration environment. US Classification 707/769, 706/47, 707/772, 707/770, 705/3, 707/784, 707/709, 705/2, 706/50, 707/783; International Classification G06F17/30, G06F21/30; Cooperative Classification G06F21/30, October 2014 <http://www.google.com/patents/US8856169>.

[95] Hu Y, Hosain MA, Jain T, Gunapati YR, Elkin L, Zhang G, et al. Global SunFarm data acquisition network, energy CRADLE, and time series analysis. In: 2013 IEEE Energytech. Cleveland (OH); 2013. p. 1–5, http://dx.doi.org/10.1109/EnergyTech.2013.6645317.

[96] hadoop.apache.org, Hadoop; 2014 <http://hadoop.apache.org>.

[97] hbase.apache.org, HBase; 2014 <http://hbase.apache.org>.

[98] Gunapati VY. Cloud based distributed computing platform for multimodal energy data streams, Master of science thesis. Case Western Reserve University, 00000; 2014 <https://etd.ohiolink.edu/ap/10?108894103917738::NO:10:P10_ETD_SUBID:94769>.

[99] Cloudera I. Cloudera express, 00000 accessed November 2014; 2014 <http://www.cloudera.com/content/cloudera/en/products-and-services/cloudera-express.html>.

[100] RHadoop; 2014 <https://github.com/RevolutionAnalytics/RHadoop/wiki>.

[101] Kline RB. Principles and practice of structural equation modeling. Guilford Press; 2011.

[102] Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bullet 1988;103(3):411–23. http://dx.doi.org/10.1037/0033-2909.103.3.411. <http://psycnet.apa.org/journals/bul/103/3/411.html>.

[103] Bollen KA, Harden JJ, Ray S, Zavisca J. BIC and alternative bayesian information criteria in the selection of structural equation models. Struct Equat Model: A Multidiscip J 2014;21(1):1–19. http://dx.doi.org/10.1080/10705511.2014.856691. http://dx.doi.org/10.1080/10705511.2014.856691.

[104] MacCallum RC, Austin JT. Applications of structural equation modeling in psychological research. Ann Rev Psychol 2000;51(1):201–26. http://dx.doi.org/10.1146/annurev.psych.51.1.201. http://dx.doi.org/10.1146/annurev.psych.51.1.201.

[105] Baumgartner H, Homburg C. Applications of structural equation modeling in marketing and consumer research: a review. Int J Res Market 1996;13(2):139–61. http://dx.doi.org/10.1016/0167-8116(95)00038-0. <http://www.sciencedirect.com/science/article/pii/0167811695000380>.

[106] Kreyssig P, Escuela G, Reynaert B, Veloz T, Ibrahim B, Dittrich P. Cycles and the qualitative evolution of chemical systems. PLoS ONE 2012;7(10):e45772. http://dx.doi.org/10.1371/journal.pone.0045772. http://dx.doi.org/10.1371/journal.pone.0045772.

[107] Bruckman LS, Wheeler NR, Kidd IV, Sun J, French RH. Photovoltaic lifetime and degradation science statistical pathway development: acrylic degradation. In: SPIE Solar Energy+ Technology, vol. 8825. International Society for Optics and Photonics; 2013. p. 88250D-8, http://dx.doi.org/10.1117/12.2024717 <http://dx.doi.org/10.1117/12.2024717>.

[108] Schuetz MA, Shell KA, Brown SA, Reinbolt GS, French RH, Davis RJ. Design and construction of a 7× low-concentration photovoltaic system based on compound parabolic concentrators. IEEE J Photovolt 2012;2(3):382–6. http://dx.doi.org/10.1109/JPHOTOV.2012.2186283.

[109] Miller DC, Kurtz SR. Durability of Fresnel lenses: a review specific to the concentrating photovoltaic application. Solar Energy Mater Solar Cells 2011;95(8):2037–68. <http://journals.ohiolink.edu/ejc/article.cgi?issn=09270248&issue=v95i0008&article=2037_doflarsttcpa>.

[110] Mahoney A, Cannon J, Woodworth J. Accelerated UV-aging of acrylic materials used in PV concentrator systems. In: Conference record of the twenty third IEEE photovoltaic specialists conference, 1993; 1993. p. 1216 – 21, http://dx.doi.org/10.1109/PVSC.1993.346948.

[111] Day M, Wiles DM. Photochemical decomposition mechanism of poly(ethylene terephthalate). J Poly Sci Part B: Poly Lett 1971;9(9):665–9. http://dx.doi.org/10.1002/pol.1971.110090906. <http://onlinelibrary.wiley.com/doi/10.1002/pol.1971.110090906/abstract.

[112] Day M, Wiles DM. Photochemical degradation of poly(ethylene terephthalate). i. Irradiation experiments with the xenon and carbon arc. J Appl Poly Sci 1972;16(1). http://dx.doi.org/10.1002/app.1972.070160116. <http://onlinelibrary.wiley.com/doi/10.1002/app.1972.070160116/abstract>.

[113] Day M, Wiles DM. Photochemical degradation of poly(ethylene terephthalate). III. Determination of decomposition products and reaction mechanism. J Appl Poly Sci 1972;16(1). http://dx.doi.org/10.1002/app.1972.070160118. <http://onlinelibrary.wiley.com/doi/10.1002/app.1972.070160118/abstract>.

[114] Day M, Wiles DM. Photochemical degradation of poly(ethylene terephthalate). II. Effect of wavelength and environment on the decomposition process. J Appl Poly Sci 1972;16(1). http://dx.doi.org/10.1002/app.1972.070160117. <http://onlinelibrary.wiley.com/doi/10.1002/app.1972.070160117/abstract>.

[115] McMahon W, Birdsall HA, Johnson GR, Camilli CT. Degradation studies of polyethylene terephthalate. J Chem Eng Data 1959;4(1):57–79. http://dx.doi.org/10.1021/je60001a009. <http://pubs.acs.org/cgi-bin/doilookup/?10.1021/je60001a009>.

[116] Buxbaum LH. The degradation of poly(ethylene terephthalate). Angew Chem Int Ed Eng 1968;7(3):182–90. http://dx.doi.org/10.1002/anie.196801821. <http://onlinelibrary.wiley.com/doi/10.1002/anie.196801821/abstract>.

[117] Launay A, Thominette F, Verdu J. Hydrolysis of poly(ethylene terephthalate): a kinetic study. Poly Degrad Stab 1994;46(3):319–24. http://dx.doi.org/10.1016/0141-3910(94)90148-1. <http://www.sciencedirect.com/science/article/pii/0141391094901481>.

[118] Bellenger V, Ganem M, Mortaigne B, Verdu J. Lifetime prediction in the hydrolytic ageing of polyesters. Poly Degrad Stab 1995;49(1):91–7. http://dx.doi.org/10.1016/0141-3910(95)00049-R. <http://www.sciencedirect.com/science/article/pii/014139109500049R>.

[119] Burgoyne CJ, Merii AL. On the hydrolytic stability of polyester yarns. J Mater Sci 2007;42(8):2867–78. http://dx.doi.org/10.1007/s10853-006-1360-5. <http://link.springer.com/article/10.1007/s10853-006-1360-5>.

[120] Turnbull L, Liggat JJ, MacDonald WA. Ageing of poly(ethylene terephthalate) and poly(ethylene naphthalate) under moderately accelerated conditions. J Appl Poly Sci 2012;124(6). http://dx.doi.org/10.1002/app.35476. <http://doi.wiley.com/10.1002/app.35476>.

[121] Pickett JE, Coyle DJ. Hydrolysis kinetics of condensation polymers under humidity aging conditions. Poly Degrad Stab 2013;98(7):1311–20. http://dx.doi.org/10.1016/j.polymdegradstab.2013.04.001. <http://linkinghub.elsevier.com/retrieve/pii/S0141391013000918>.

[122] Oreski G, Wallner G. Aging mechanisms of polymeric films for PV encapsulation. Solar Energy 2005;79(6):612–7. http://dx.doi.org/10.1016/j.solener.2005.02.008. <http://linkinghub.elsevier.com/retrieve/pii/S0038092X05000794>.

[123] De Soto W, Klein SA, Beckman WA. Improvement and validation of a model for photovoltaic array performance. Solar Energy 2006;80(1):78–88, 0046. http://dx.doi.org/10.1016/j.solener.2005.06.010. <http://www.sciencedirect.com/science/article/pii/S0038092X05002410>.

[124] Pysch D, Mette A, Glunz S. A review and comparison of different methods to determine the series resistance of solar cells. Solar Energy Mater Solar Cells 2007;91(18):1698–706. http://dx.doi.org/10.1016/j.solmat.2007.05.026. <http://www.sciencedirect.com/science/article/pii/S0927024807002255>.

[125] Yi Y, Peshek TJ, French RH, Shan J, Singer KD. Optical and photochemical effects of natural nanoparticles on screen printed silver conductive lines of a commercial photovoltaic cell. J Appl Phys, submitted for publication.

[126] Czanderna A, Pern F. Encapsulation of PV modules using ethylene vinyl acetate copolymer as a pottant: A critical review. Solar Energy Mater Solar Cells 1996;43(2):101–81. http://dx.doi.org/10.1016/0927-0248(95)00150-6. <http://www.sciencedirect.com/science/article/pii/0927024895001506>.

[127] Hossain MA, Xu Y, Peshek TJ, Ji L, Sun J, Abramson AR, et al. Microinverter thermal performance in the real-world: Measurements and modeling. PLoS One, submitted for publication.

[128] Borgman CL. The conundrum of sharing research data. J Am Soc Inf Sci Technol 2012;63(6):1059–78. http://dx.doi.org/10.1002/asi.22634. <http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/abstract>.

[129] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet 2000;25(1):25–9. http://dx.doi.org/10.1038/75556. <http://www.nature.com/ng/journal/v25/n1/abs/ng0500_25.html>.

[130] Zhang G-Q, Sahoo SS, Lhatoo SD. From classification to epilepsy ontology and informatics. Epilepsia 2012;53:28–32. http://dx.doi.org/10.1111/j.1528-1167.2012.03556.x. <http://onlinelibrary.wiley.com/doi/10.1111/j.1528-1167.2012.03556.x/abstract>.

[131] Unified medical language system (UMLS) – home; 2014 <http://www.nlm.nih.gov/research/umls/>.

[132] Zhang G-Q, Cui L, Lahtoo SD, Sahoo SS. MEDCIS: multi-modality epilepsy data capture and integration system | AMIA knowledge center. In: AMIA annual symposium of proceedings; 2014. p. 1248–57 <http://knowledge.amia.org/56638-amia-1.1540970/t-004-1.1544972/f-004-1.1544973/a-228-1.1544989/an-228-1.1544990?qr=1>.

[133] Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al. Big data: the next frontier for innovation, competition, and productivity, Tech. rep., May 2011 <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>.

[134] Califano A. An assessment of the impact of the NCI cancer biomedical informatics grid (caBIG), Tech. rep. National Cancer Institute, March 2011.

[135] Sleep data – national sleep research resource – NSRR; 2014 <https://sleepdata.org/>.

[136] Center for SUDEP research, 00000 2014 <http://csr.case.edu/csr/index.php/Main_Page>.

[137] Cui L. Ontology-guided health information extraction, organization, and exploration, Ph.D. thesis. Case Western Reserve University, 00001; 2014 <https://etd.ohiolink.edu/ap/10?0::NO:10:P10_ETD_SUBID:95790>.

[138] Tao GQ, Cui L, Zhu W, Sun M, Bodenreider O, Zhang GQ. Mining relation reversals in the evolution of SNOMED CT. San Francisco (CA); 2015 [in press] <http://www.amia.org/jointsummits2015>.

[139] Zhang GQ, Zhu W, Sun M, Tao S, Bodenreider O, Cui L. MaPLE: a MapReduce pipeline for lattice-based evaluation and its application to SNOMED CT,. Washington (DC); 2014. pp. 754–9 [00000].

[140] Hu Y, Xu Y, Peshek TJ, Du W, Sun J, French RH. A data science procedure for forecasting brand-dependent regression model to predict photovoltaic performance.

[141] Faraway J, Mahabal A, Sun J, Wang X, Zhang L. et al. Modeling light curves for improved classification, arXiv preprint arXiv:1401.3211 <http://arxiv.org/abs/1401.3211>.

[142] Xu Y, Sun J, Carter RR, Bogie KM. Personalized prediction of chronic wound healing: an exponential mixed effects model using stereophotogrammetric measurement. J Tissue Viab 2014;23(2):48–59. http://dx.doi.org/10.1016/j.jtv.2014.04.001. <http://www.sciencedirect.com/science/article/pii/S0965206X1400028X>.

[143] Ganocy SJ, Sun J. Heteroscedastic change point analysis and application to footprint data. J Data Sci, in press.

[144] Ganocy S, Sun J, Wang Y. hcp: change point estimation for regression with heteroscedastic data; 2014 <http://cran.r-project.org/web/packages/hcp/>.

[145] Warner RM. Spectral analysis of time-series data. Methodology in the social sciences. The Guilford Press; 1998. <http://www.guilford.com/books/Spectral-Analysis-of-Time-Series-Data/Rebecca-Warner/9781572303386>.

[146] Einstein A. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. Annalen der Physik 1905;17:549–69.

[147] Van Kampen N. Stochastic processes in physics and chemistry. North-Holland personal library.

[148] Glansdorff P, Prigogine I. Thermodynamic theory of structure, stability and fluctuations. New York: Wiley Interscience; 1971.

[149] Hori Y, Hara S. Noise-induced spatial pattern formation in stochastic reaction–diffusion systems. In: 2012 IEEE 51st annual conference on decision and control (CDC); 2012. p. 1053–58, http://dx.doi.org/10.1109/CDC.2012.6426152.

[150] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377–81. http://dx.doi.org/10.1016/j.jbi.2008.08.010. <http://linkinghub.elsevier.com/retrieve/pii/S1532046408001226>.