Primer on Analysis of Experimental Data and Design of Experiments

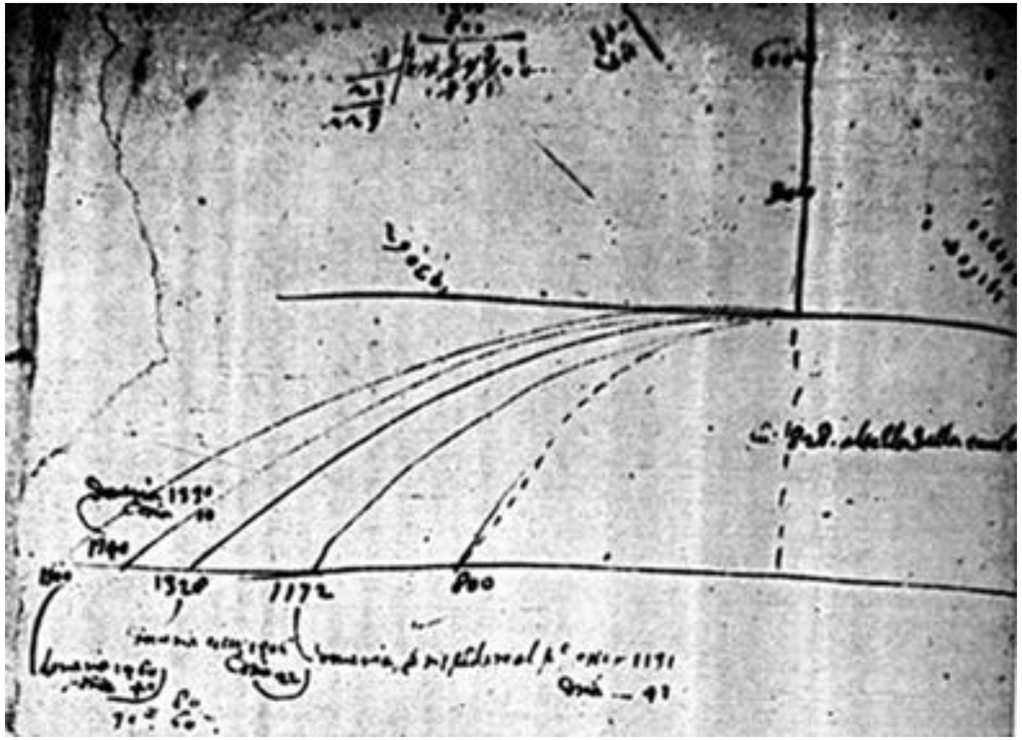
Lecture 14. Physics-based Machine Learning

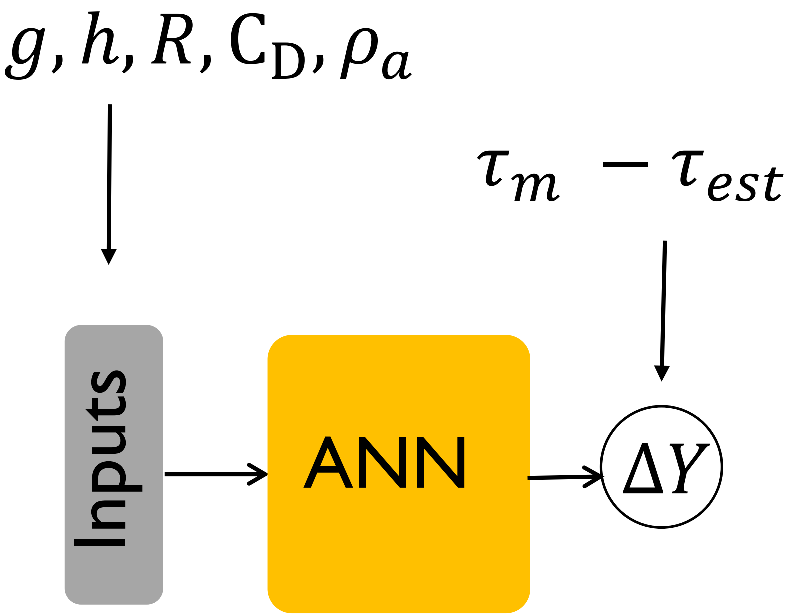
Muhammad A. Alam
alam@purdue.edu

Outline

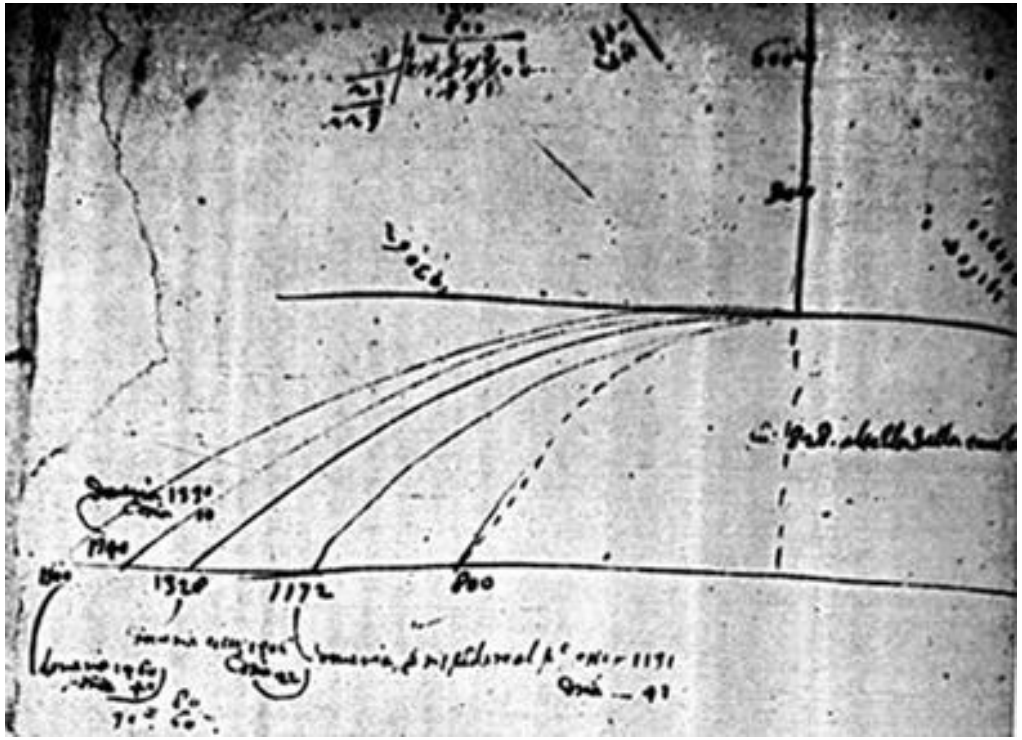
1. Why and what of physics-based machine learning
2. Example 1: Dropping a ball in the real world
3. Example 2: Lake temperature distribution
4. Approach 2: Structural Equation Modeling
2. Conclusions

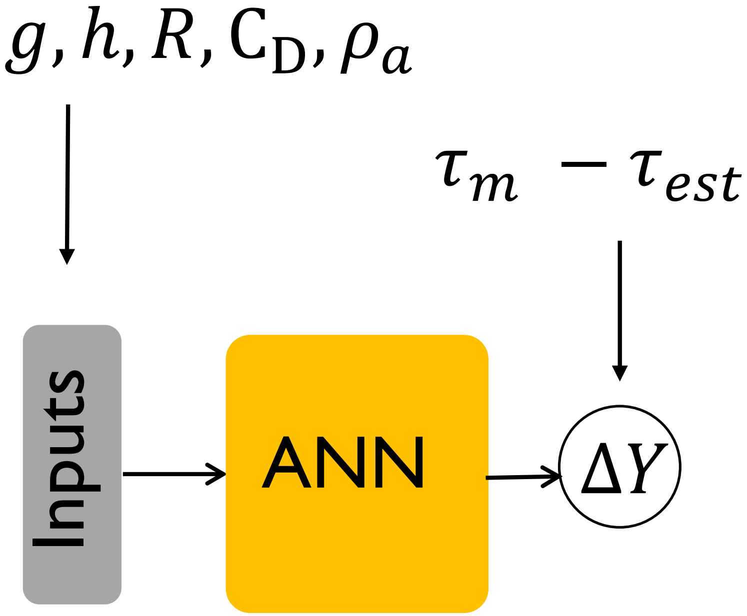
Galileo experiments





Galileo vs. Newton

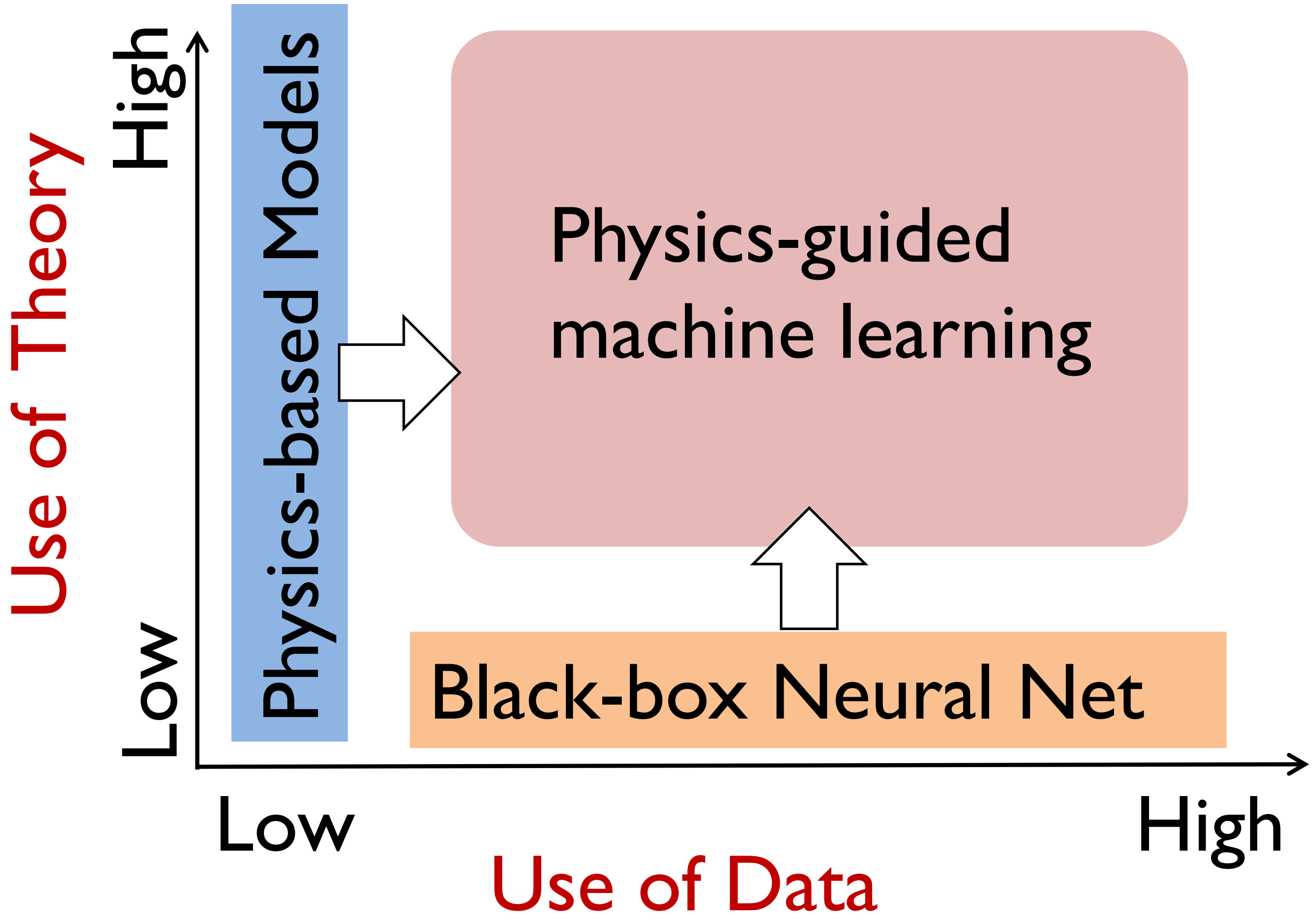




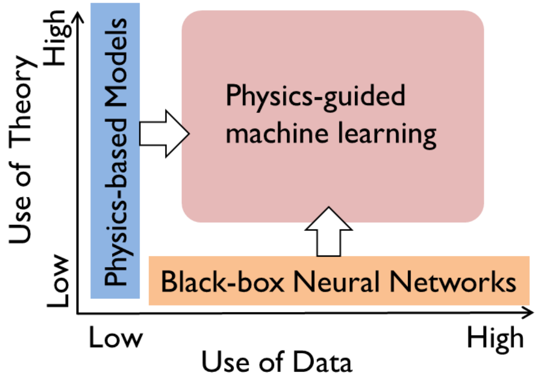
$$\frac{d^2 z}{dt^2} = -\frac{g}{m}$$

$$\tau_0 = \sqrt{\frac{2h}{g}}$$

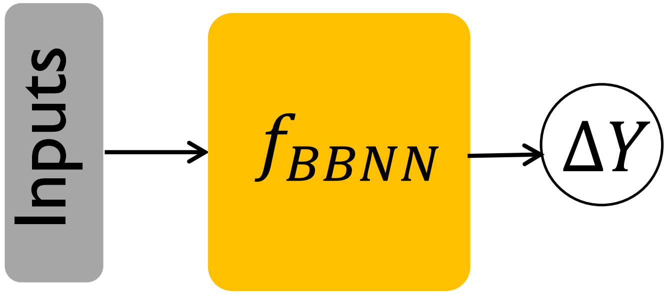
Physics-based machine learning approach



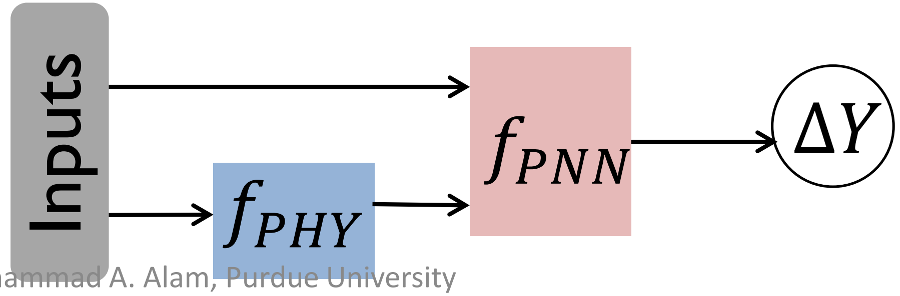
Physics-based machine learning approach



Current approach



An improved approach



Outline

1. Why and what of physics-based machine learning

2. Example 1: Dropping a ball in the real world

3. Example 2: Lake temperature distribution

4. Example 3:

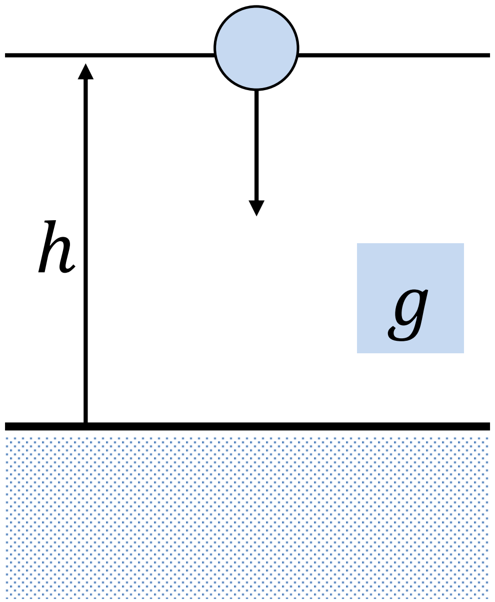
2. Conclusions

A ball falling under gravity (idealized)

$$\frac{d^2z}{dt^2} = -g \Rightarrow z = h - \frac{gt^2}{2}$$

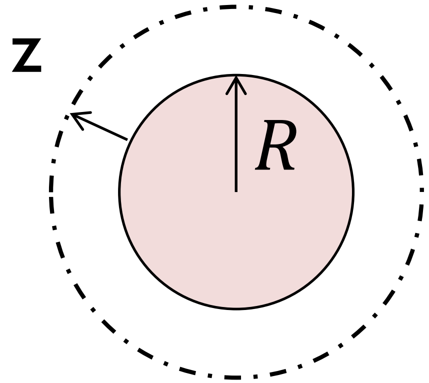
$$\Rightarrow h = \frac{g\tau_0^2}{2}$$

$$\tau_0 = \sqrt{\frac{2h}{g}}$$



The actual ball will experience height dependent gravity, air resistance that depends on temperature and humidity, etc.

A falling ball with height dependent gravity (idealized)



$$\frac{d^2z}{dt^2} = -\frac{g}{(1+z/R)^2}$$

$$\tau_1 = \tau_0 \cosh^{-1}(1 - 2x) \qquad x = z/R$$

$$\tau_1 \sim \tau_0 \left(1 + \frac{5}{6}x + \frac{43}{40}x^2 + \frac{177}{112}x^3 + \ldots\right)$$

C. F. Bohren, "Dimensional analysis, falling bodies, and the fine art of not solving differential equations,", 2003.

A falling ball with air resistance

$$m\frac{dv}{dt} = -mg + ½\,\rho_a\,(\pi r^2)C_D\,v^2 \rightarrow \frac{dv}{dt} = -g + bv^2$$

ρ_a ... atmospheric density
C_D … drag coefficient
R …. Radius of the ball

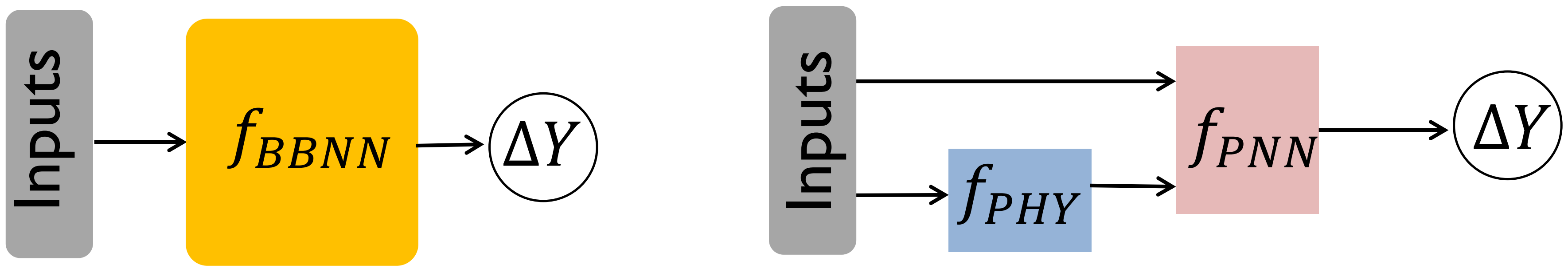
$$\cosh\left(t\sqrt{gb}\right) = \exp(hb) \qquad z = h - \frac{\ln\cosh\left(t\sqrt{gb}\right)}{b}$$

$$v = \frac{dz}{dt} = \sqrt{\frac{g}{b}}\tanh\left(t\sqrt{gb}\right)$$

A falling ball with gravity and resistance

$$\frac{dv}{dt} = -\frac{g}{(1+z/R)^2} + bv^2$$

A numerical/perturbative solution may not be possible. In practice, b(z) is unknown function of humidity, temperature, etc. A machine learning approach is preferred.



Outline

1. Why and what of physics-based machine learning

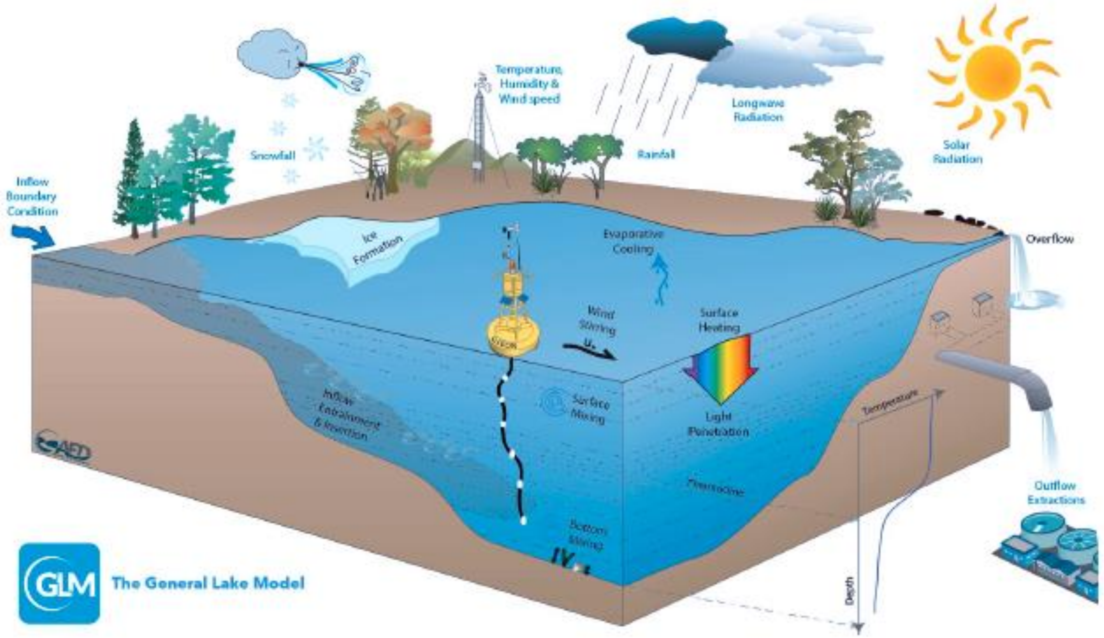
2. Example 1: Dropping a ball in the real world

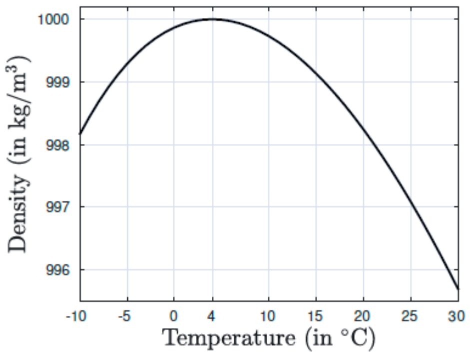
3. Example 2: Lake temperature distribution

4. Approach 2: Structural Equation Modeling

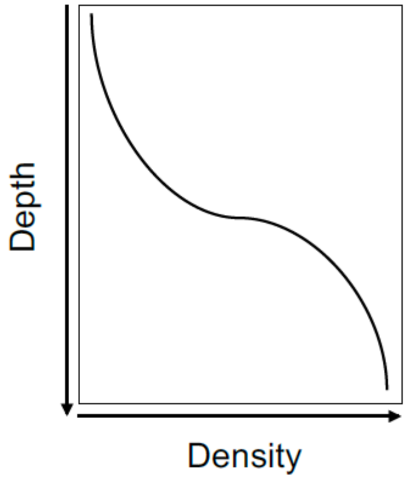
2. Conclusions

A example involving lake temperature





(a) Temperature–Density Relationship

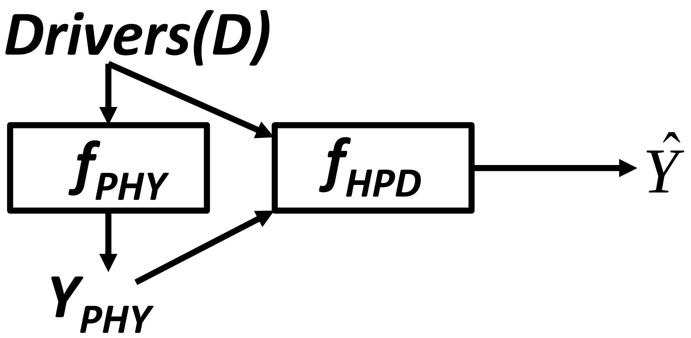


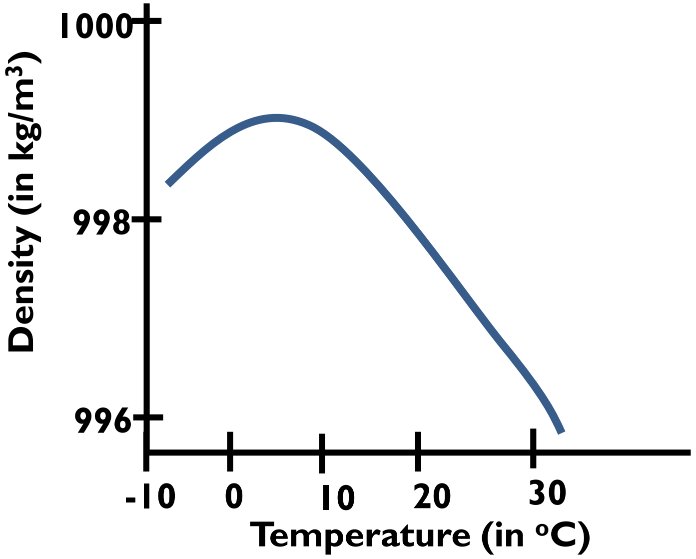
(b) Density–Depth Relationship

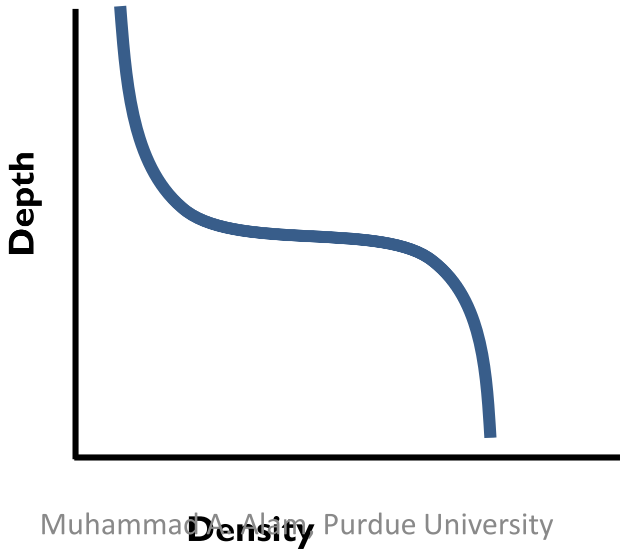
Physics-guided Neural Networks (PGNN): An application in Lake temperature Modeling Anuj Karpatne, 2016.

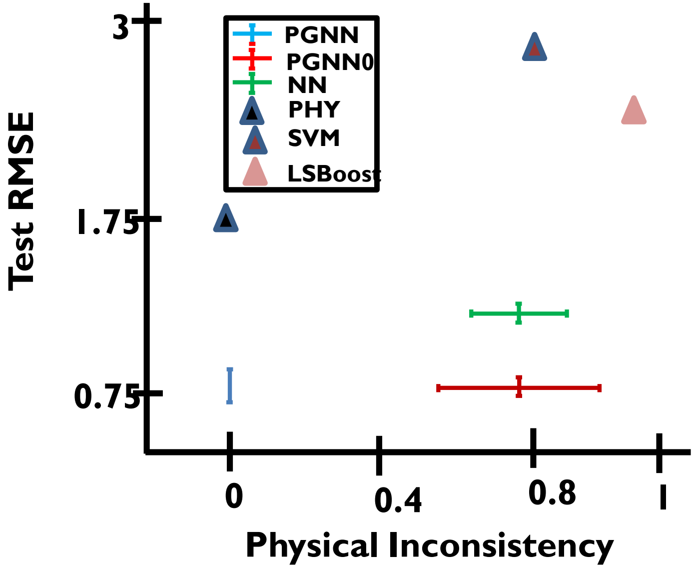
A example involving lake temperature

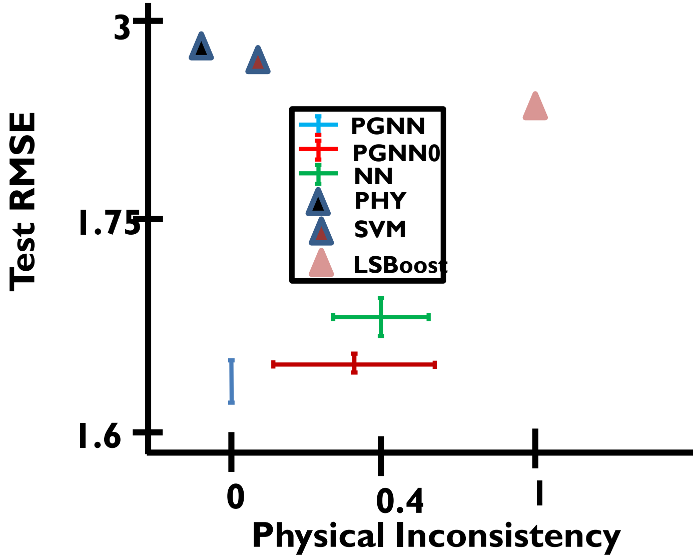
Input drivers
Day of Year (1 … 366)
Depth (m)
Short-wave Radiation (W/m^2)
Long-wave Radiation (W/m^2)
Air Temperature (Degree C)
Relative Humidity (0 … 100%)
Wind Speed (m/s)
Rain (cm)
Growing degree Days [14]
Is Freezing (True/False)
Is Snowing (True/False)











Outline

1. Why and what of physics-based machine learning

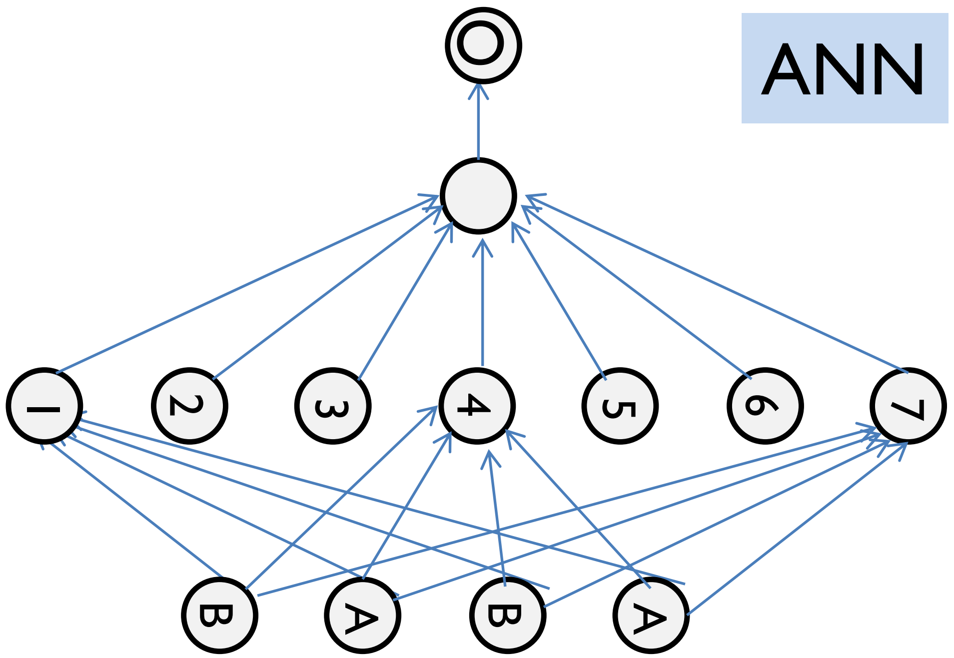
2. Example 1: Dropping a ball in the real world

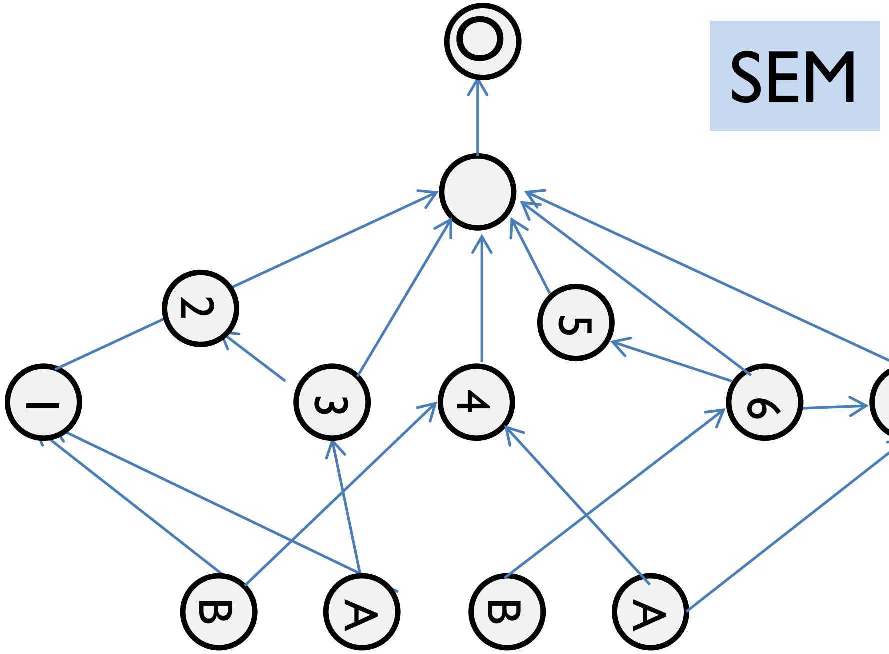
3. Example 2: Lake temperature distribution

4. Approach 2: Structural Equation Modeling

2. Conclusions

Structural Equation modeling: Motivation

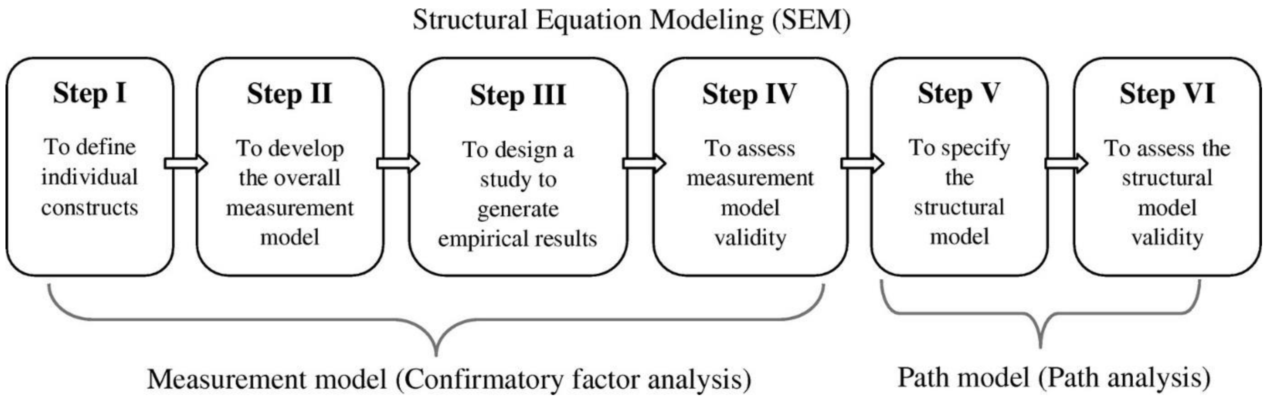


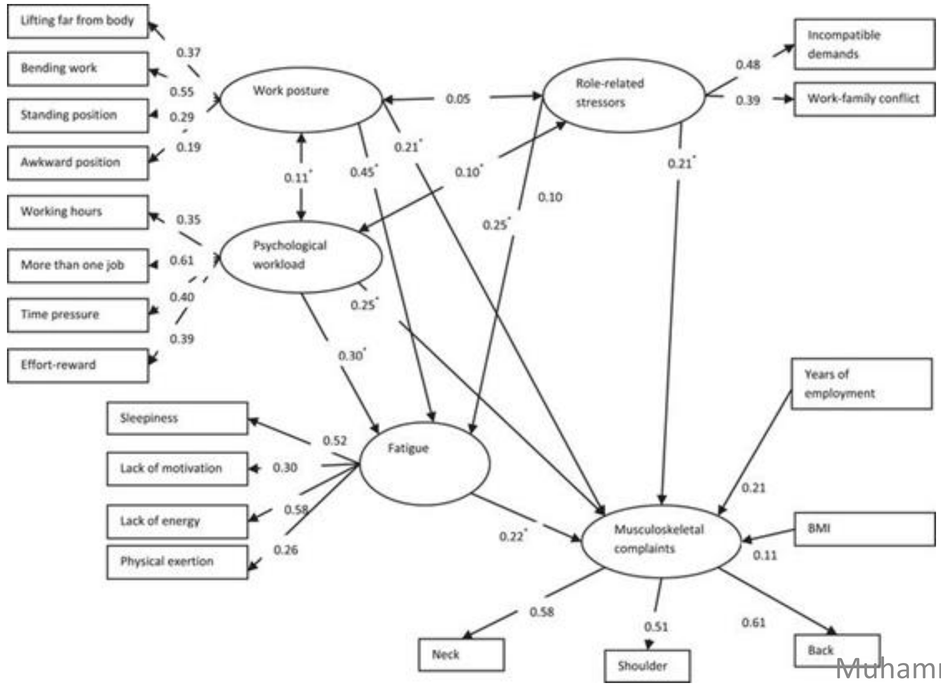


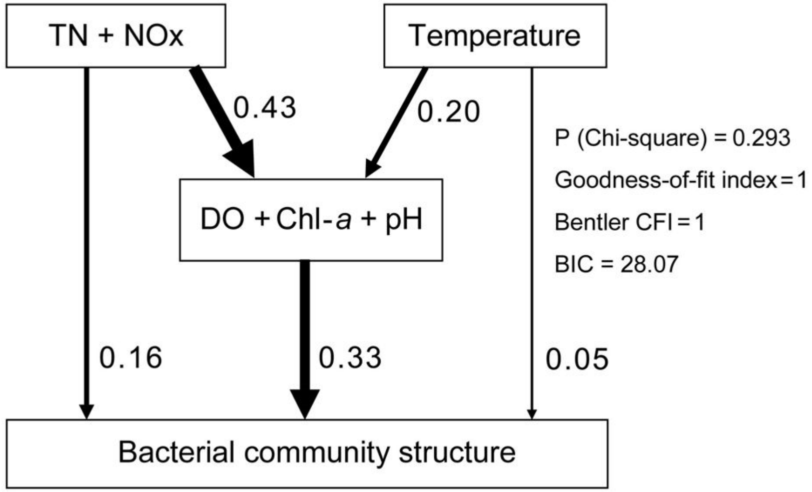
Nodes defined statistically
not appropriate for extrapolation

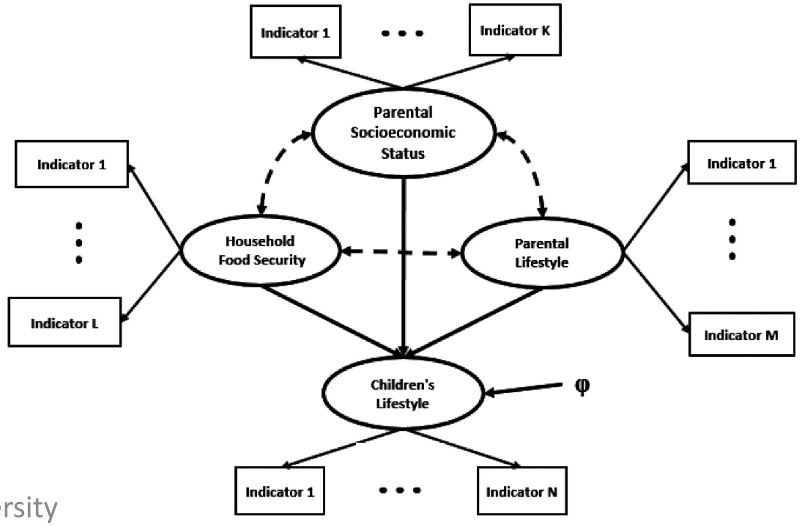
Nodes defined physically
Interconnects are nonlinear
Extrapolation possible

Structural Equation modeling: Examples







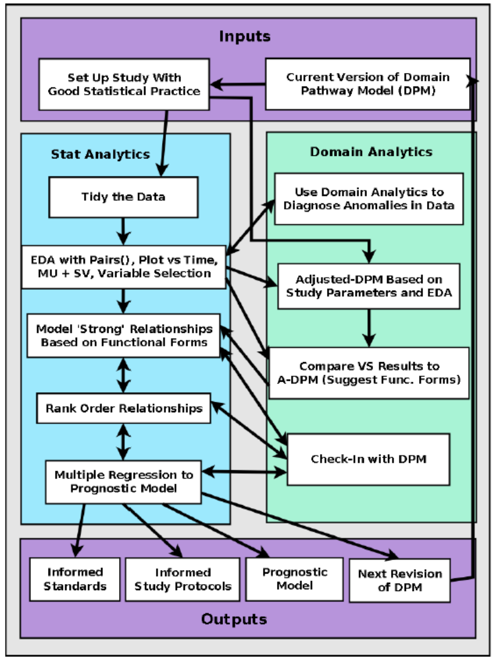


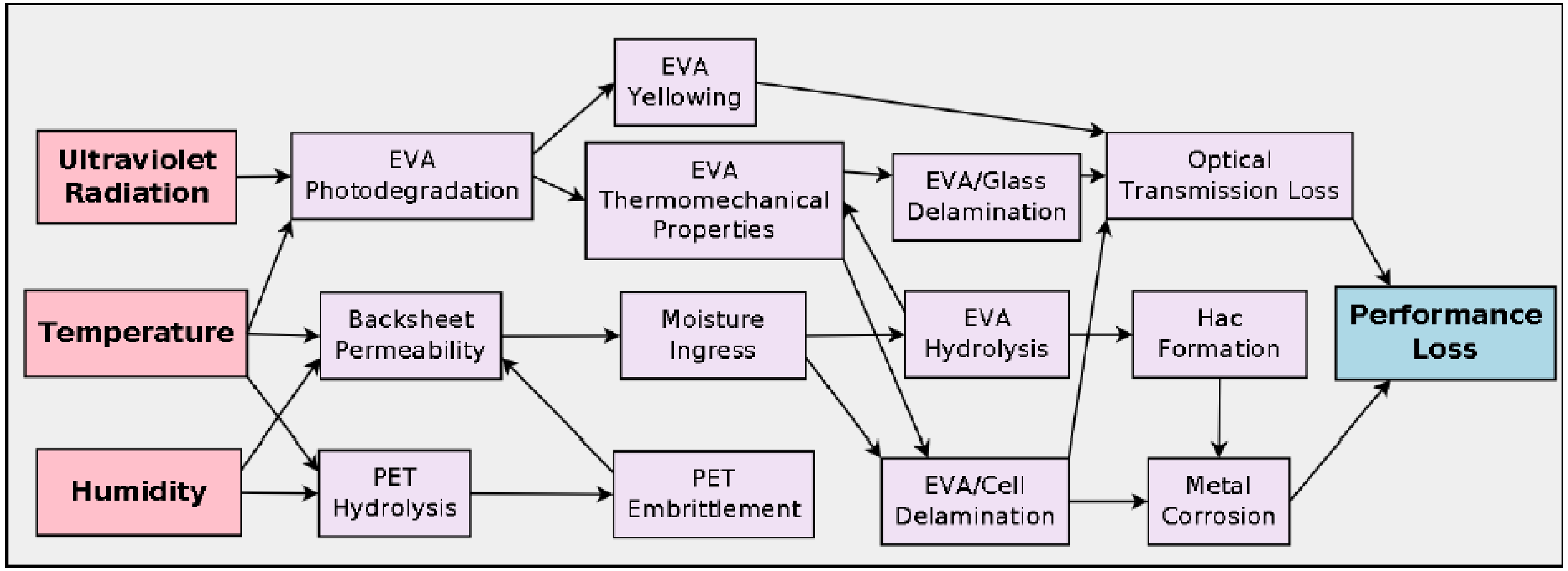
Example: Degradation of Solar Cells

Received April 16, 2013, accepted April 23, 2013, date of publication June 10, 2013, date of current version June 25, 2013.
Digital Object Identifier 10.1109/ACCESS.2013.2267611

Statistical and Domain Analytics Applied to PV Module Lifetime and Degradation Science

LAURA S. BRUCKMAN[1], NICHOLAS R. WHEELER[2], JUNHENG MA[3], ETHAN WANG[4], CARL K. WANG[4], IVAN CHOU[5], JIAYANG SUN[3], AND ROGER H. FRENCH[6] (Member, IEEE)





Variable Fits	Functional Form
Simple Linear (A)	$y = a + bx + \varepsilon$
Quadratic (B)	$y = a + bx + c * x^2 + \varepsilon$
Simple Quadratic (C)	$y = a + c * x^2 + \varepsilon$
Exponential (D)	$y = a + d * \exp^x + \varepsilon$
Logarithmic (E)	$y = a + f * (\log(x)) + \varepsilon$
Linear Change Point (F)	$a + d * (1 \pm \exp(g(x-h)) + \varepsilon$
Nonlinearizable Exponential (G-up, H-down)	$y = a + b * x + b_1 * (x - c)_+ + \varepsilon$

PV Degradation (continued)

Physical phenomenon of IRBS1, IRBS2, Hac, IR2,

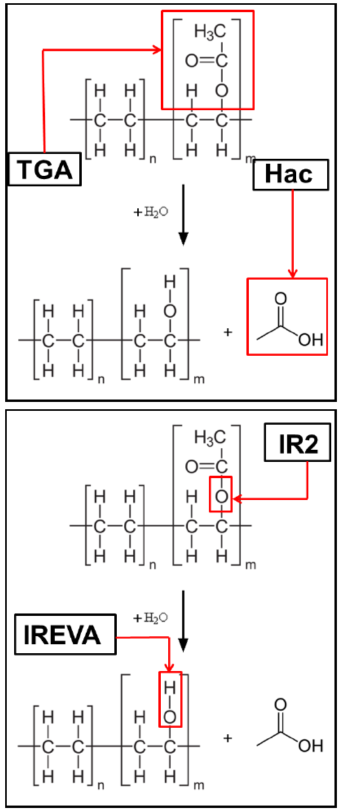


FIGURE 2. Encapsulant degradation of EVA hydrolysis. (a): The mass decrease [48] of the sample was monitored during the heating process in dependence of the temperature. The decomposition of EVA and the appearance of acetic acid can be detected. Hydrolysis of vinyl-acetate monomers in EVA results in the generation of acetic acid [49]. (b): As the exposure time increases, the acetate C=O (1735 cm^{-1}) peak decreases continuously, whereas the aldehyde/ketone C=O (1716 cm^{-1}) and O-H (near 3400 cm^{-1}) peaks increase. This results from decomposition of vinyl acetate in the EVA and the formation of aldehydes, ketones and alcohols [50]. Ester ether C-O-C /overall CH absorbance ratios were calculated to measure the relative acetate content [47].

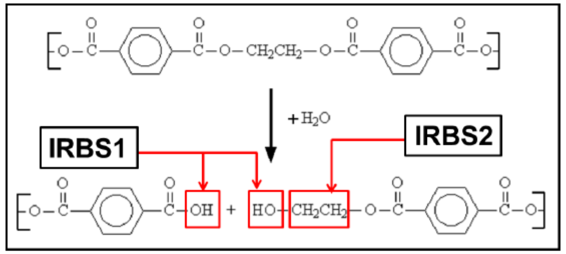


FIGURE 3. An IR peak at 3373 cm^{-1} refers to a stretching vibration of hydroxyl groups which are related to hydrolysis [51]. The changes in the stretching vibration region of methylene group (CH_2) at 2927 cm^{-1} are attributed to chain scission due to hydrolysis [52].

Variable Fits	Functional Form
Simple Linear (A)	$y = a + bx + \varepsilon$
Quadratic (B)	$y = a + bx + c * x^2 + \varepsilon$
Simple Quadratic (C)	$y = a + c * x^2 + \varepsilon$
Exponential (D)	$y = a + d * \exp^x + \varepsilon$
Logarithmic (E)	$y = a + f * (\log(x)) + \varepsilon$
Linear Change Point (F)	$a + d * (1 \pm \exp(g(x - h)) + \varepsilon$
Nonlinearizable Exponential (G-up, H-down)	$y = a + b * x + b_1 * (x - c)_+ + \varepsilon$

UV radiation

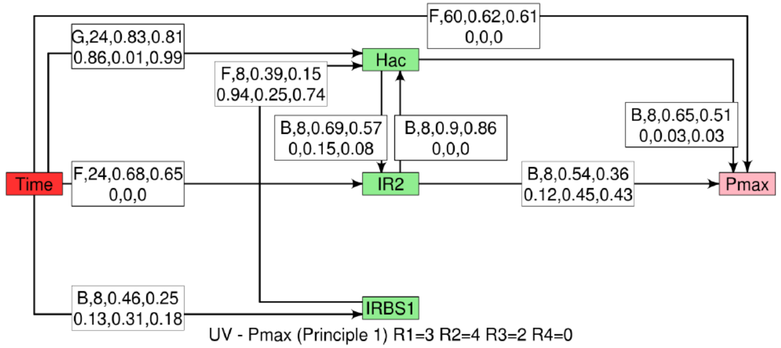


FIGURE 8. Model generated with the most relationships shown for Principle 1 for the UV exposure for the system response of *Pmax* using the unit variables of *Hac*, *IR2* and *IRBS1* shows 10 relationships. Information on each relationship is described in the box. The information contained is functional form, number of observations, R^2, adjusted R^2, P-value 1, P-value 2 and P-value 3, respectively. The strength of the SSR is summarized by the line width of the SSR border based on the R^2 value to aide visualization (below 0.2 not shown, R1 has the thinnest border (0.2-0.5), R2 (0.5-0.7), R3 (0.7-0.9) and R4 the thickest (≥ 0.9)). The functional forms are designated as A (simple linear), B (quadratic), C (simple quadratic), D (exponential), E (logarithmic), F (linear change point), G (nonlinearizable exponential-up) and H (nonlinearizable exponential-down).

Damp-heat

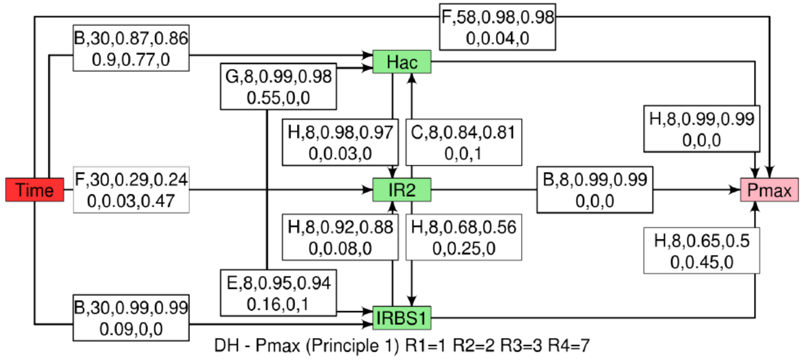
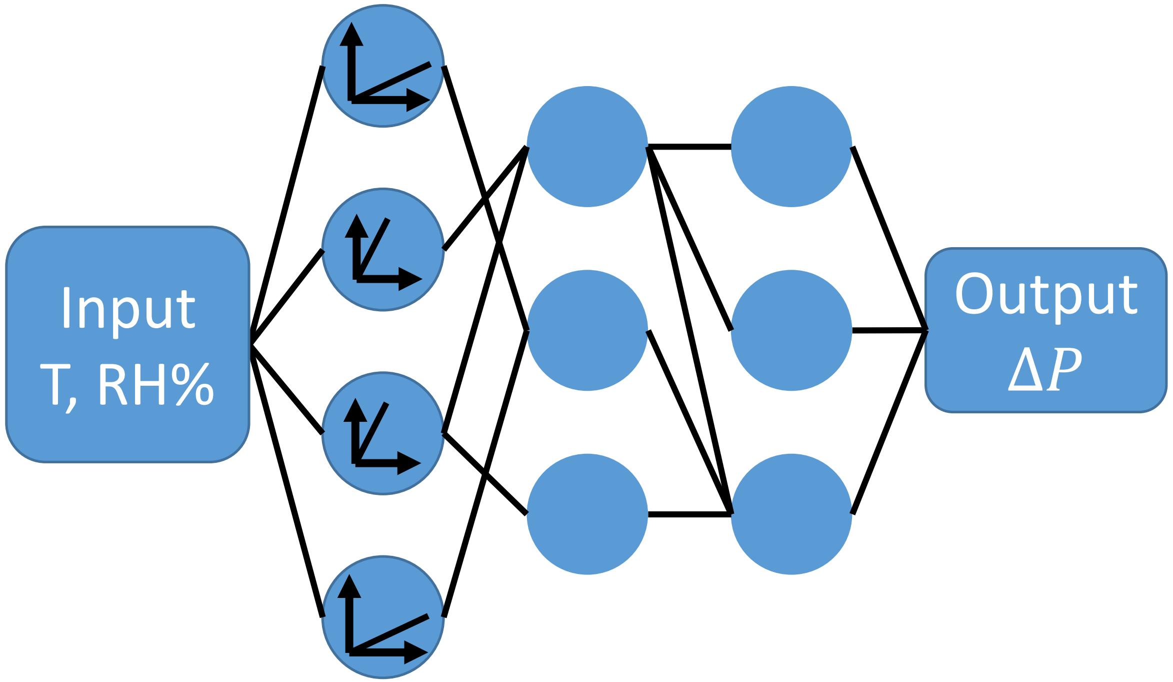
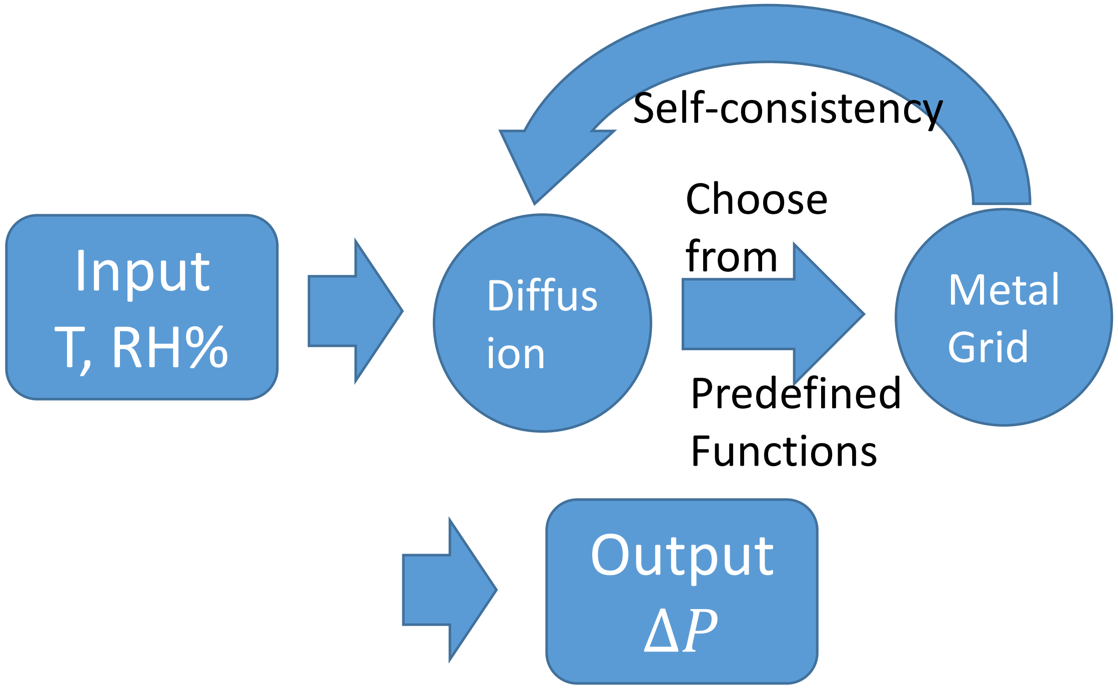


FIGURE 7. Model generated with the most relationships shown for Principle 1 for the damp heat exposure for the system response of *Pmax* using the unit variables of *Hac*, *IR2* and *IRBS1* shows 13 relationships. Information on each relationship is described in the box. The information contained is functional form, number of observations, R^2, adjusted R^2, P-value 1, P-value 2 and P-value 3, respectively. The strength of the SSR is summarized by the line width of the SSR border based on the R^2 value to aide visualization (below 0.2 not shown, R1 has the thinnest border (0.2-0.5), R2 (0.5-0.7), R3 (0.7-0.9) and R4 the thickest (≥ 0.9)). The functional forms are designated as A (simple linear), B (quadratic), C (simple quadratic), D (exponential), E (logarithmic), F (linear change point), G (nonlinearizable exponential-up) and H (nonlinearizable exponential-down).

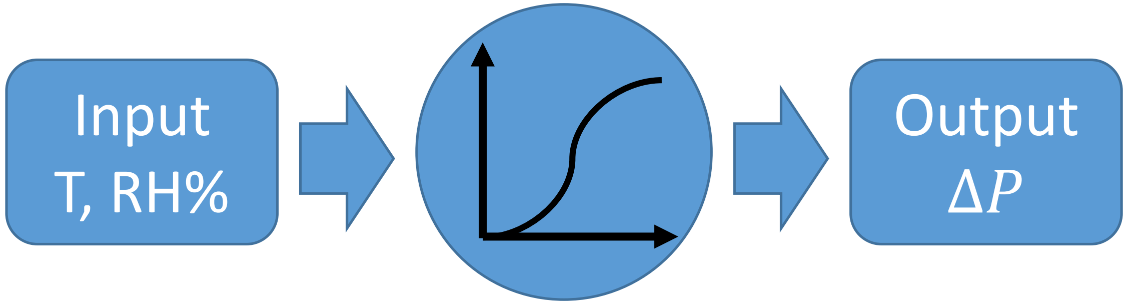
Machine Learning (Linear connection)



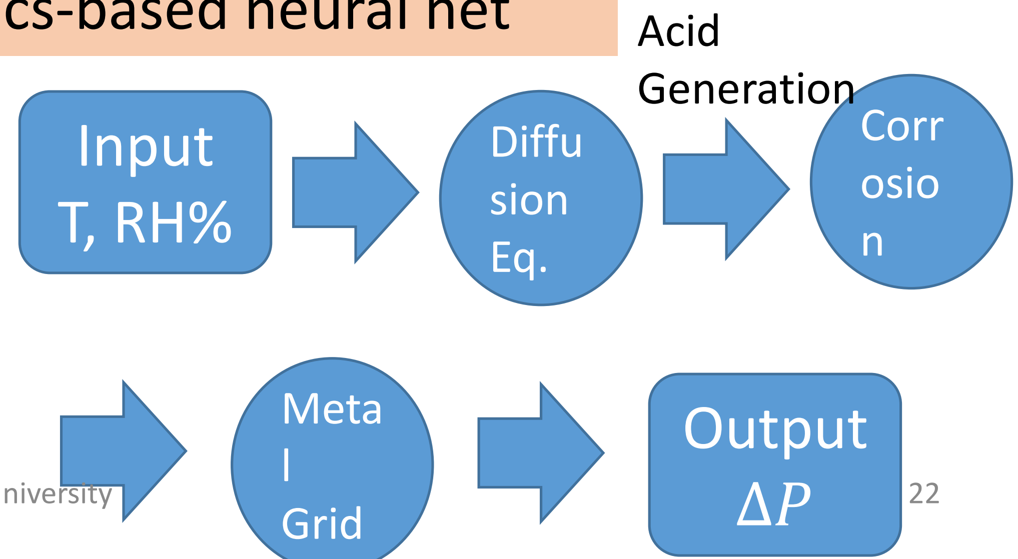
Physics-based Structural Equation Modeling



Machine Learning (Non-linear connection)



Physics-based neural net



Conclusions

1. Despite its demonstrated utility in variety of problems in commerce and business, the statistical nature of Machine Learning makes it difficult to use in engineering design or predictive modeling problems.

2. Physics-based approaches, such as physics-guided neural network or structural equation modeling offers opportunities to embed physical understanding of the model.

3. Some of the models involve non-linear coefficients – the computational efficiency of the model for very large systems are not fully understood.

4. One may be able to move from purely statistical model (e.g. Ptolemy) to physics-inspired model (e.g. Newton) one the physical understanding of the latent variable emerge over time.

Review Questions

1. Statistical machine learning is not good for out-of-range extrapolation. Explain.
2. In what ways traditional machine learning similar to a “child learning a new language or Ptolemic mode of the solar system? In what sense are the comparisons accurate?
3. Deeper the learning, shallower is the understanding. Why?
4. Name two examples where physics-based machine learning could improve predictions.
5. Give one example where physics based machine learning can actually given wrong result.
6. What is the difference between Singular value decomposition and newer machine learning algorithms?

References

Physics based Machine Learning

C. F. Bohren, "Dimensional analysis, falling bodies, and the fine art of not solving differential equations,"

Galileo experiment on projectile motion: http://galileo.rice.edu/lib/student_work/experiment95/paraintr.html

Physics-guided Neural Networks (PGNN): An application in Lake temperature Modeling Anuj Karpatne, 2016.

Structured Equation Approach.

Professor Patrick Sturgis, NCRM director, in the first (of three) part of the Structural Equation Modeling NCRM online course. Structural Equation Modeling: what is it and what can we use it for? (part 1 of 6) https://www.youtube.com/watch?v=eKkESdyMG9w

https://www.youtube.com/watch?v=-m4ag3WQcCw

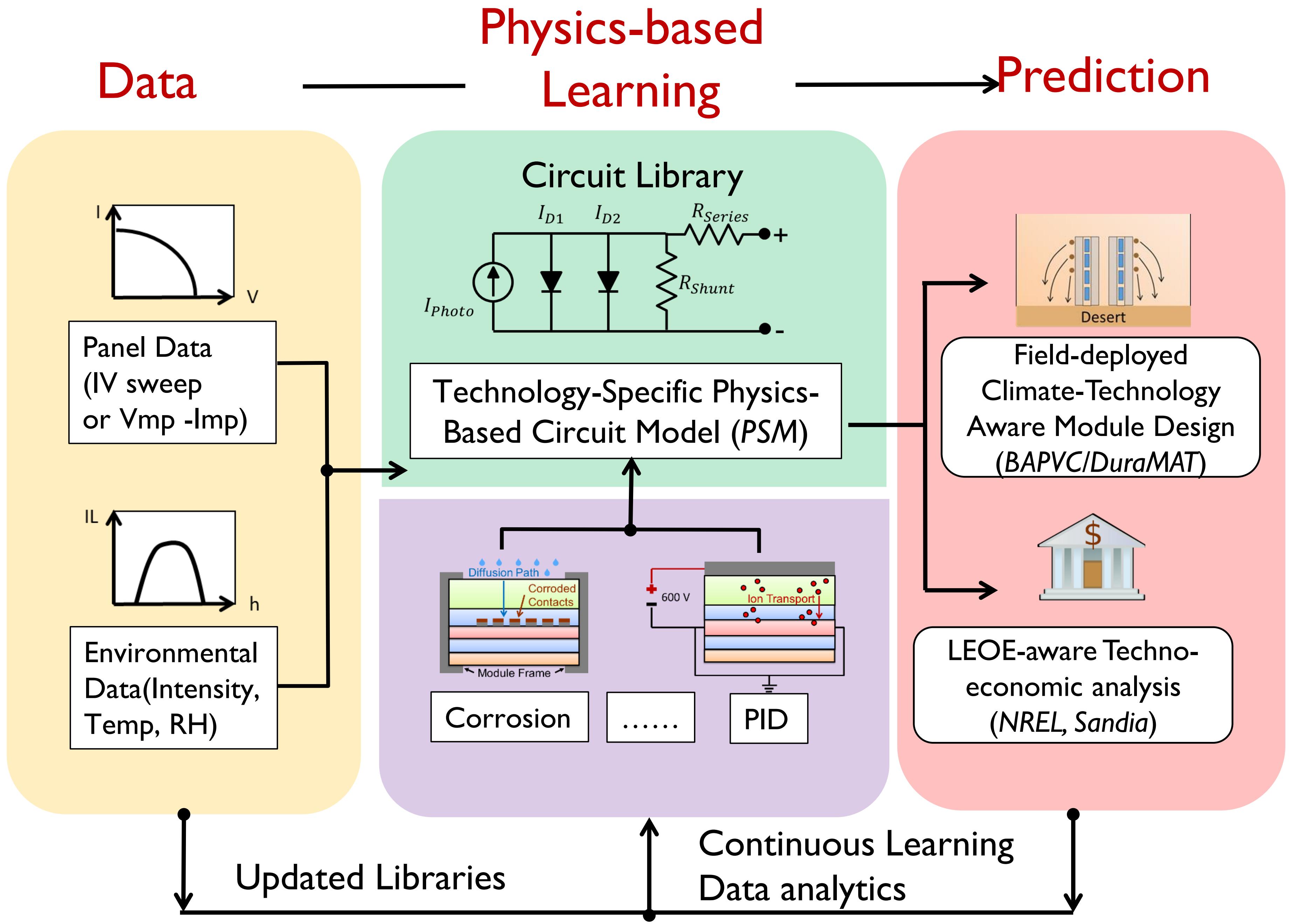
A talk about three principal advantages of structural equation models (SEMs) relative to more traditional analytic techniques, like the linear regression model. These include...

(1) The ability to represent constructs as latent variables that are uncontaminated by measurement error
(2) Falsification tests and indices of fit with to evaluate the tenability of a proposed theoretical model
(3) Flexibility to allow researchers to specify statistical models that more closely match theory

Dan describes these advantages using a specific example on the factors that relate to young children's popularity with peers.

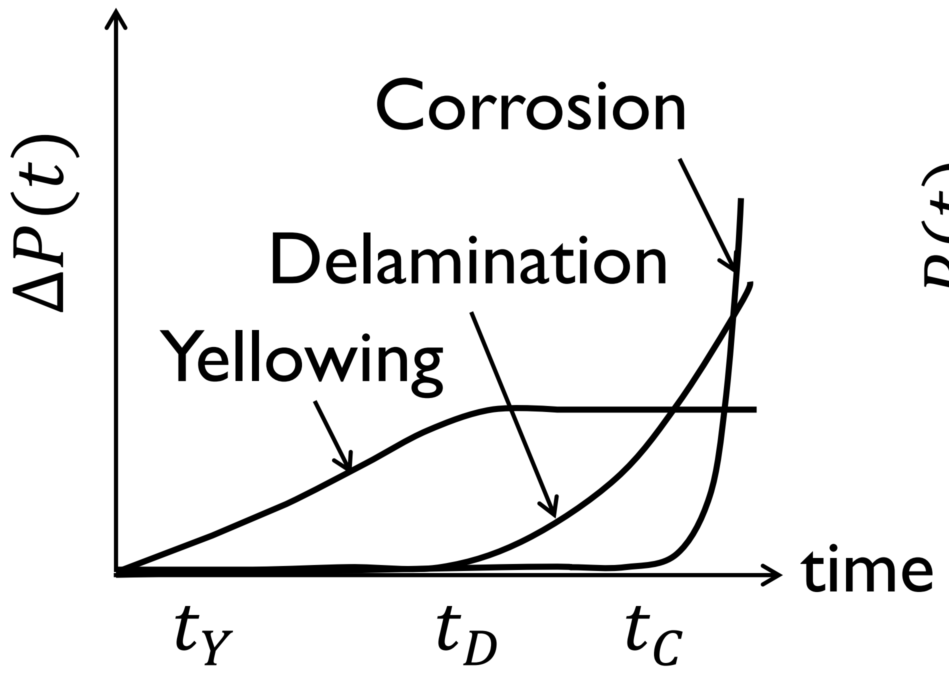
In addition to these three principal advantages of the SEM, there are many other ways that the model can be expanded and used to address interesting theoretical questions. For instance, a variety of SEMs exist for analyzing longitudinal data, including latent growth curve models and latent change score models. SEMs also provide a powerful framework within which to evaluate population heterogeneity, including differences over known groups (e.g., boys and girls) or latent groups (e.g., clusters of individuals for whom predictive relationships differ). For those interested in learning more, we offer summer training seminars on SEM and longitudinal SEM, see http://www.curranbauer.org/training/.

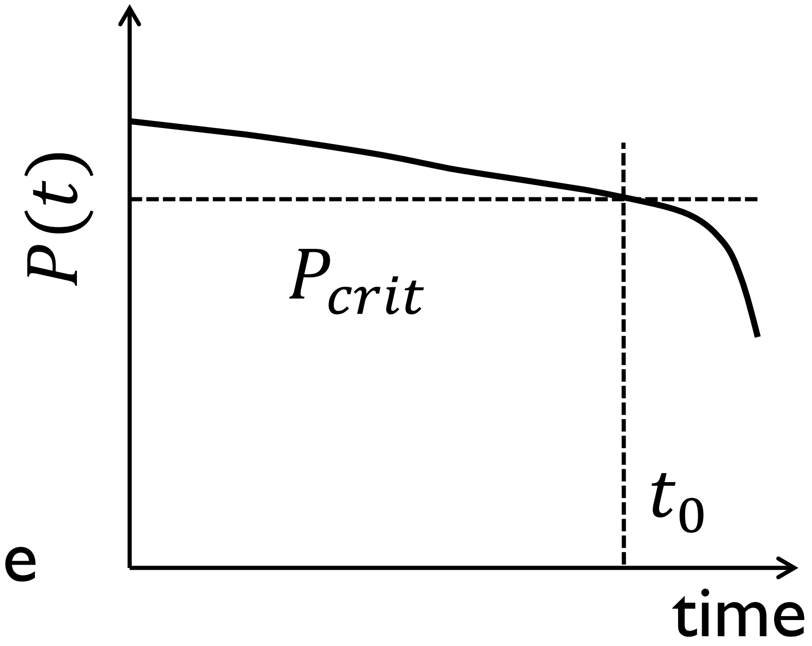
Appendix

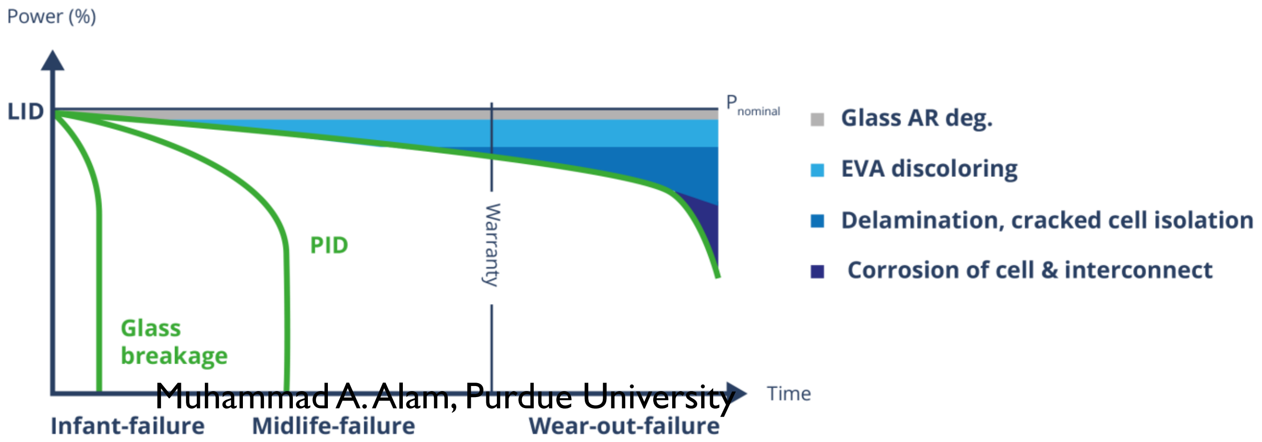


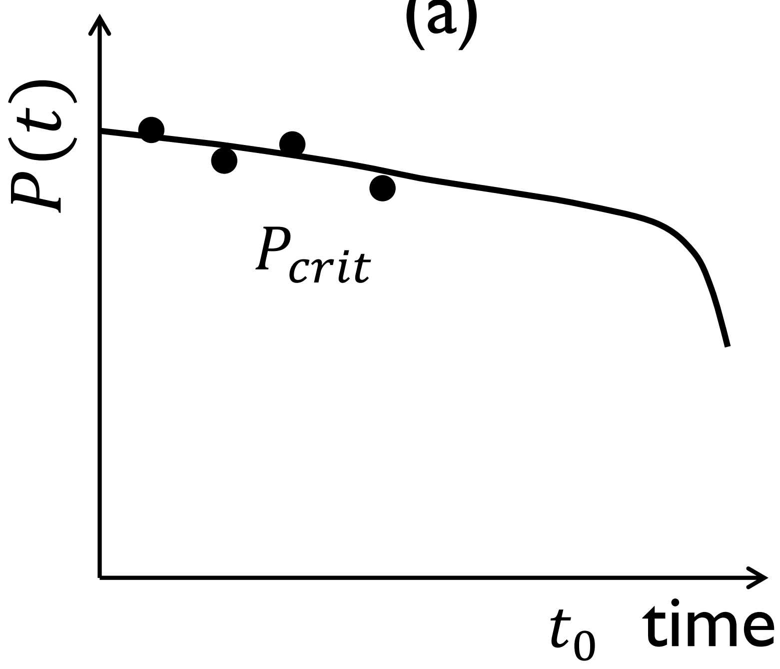
Data
Physics-based Learning
Prediction
I
V
Panel Data (IV sweep or Vmp -Imp)
IL
h
Environmental Data(Intensity, Temp, RH)
Circuit Library
I_{D1}
I_{D2}
R_{Series}
+
R_{Shunt}
I_{Photo}
-
Technology-Specific Physics-Based Circuit Model (*PSM*)
Diffusion Path
Corroded Contacts
Module Frame
600 V
Ion Transport
Corrosion
……
PID
Desert
Field-deployed Climate-Technology Aware Module Design (*BAPVC/DuraMAT*)
$
LEOE-aware Techno-economic analysis (*NREL, Sandia*)
Updated Libraries
Continuous Learning
Data analytics

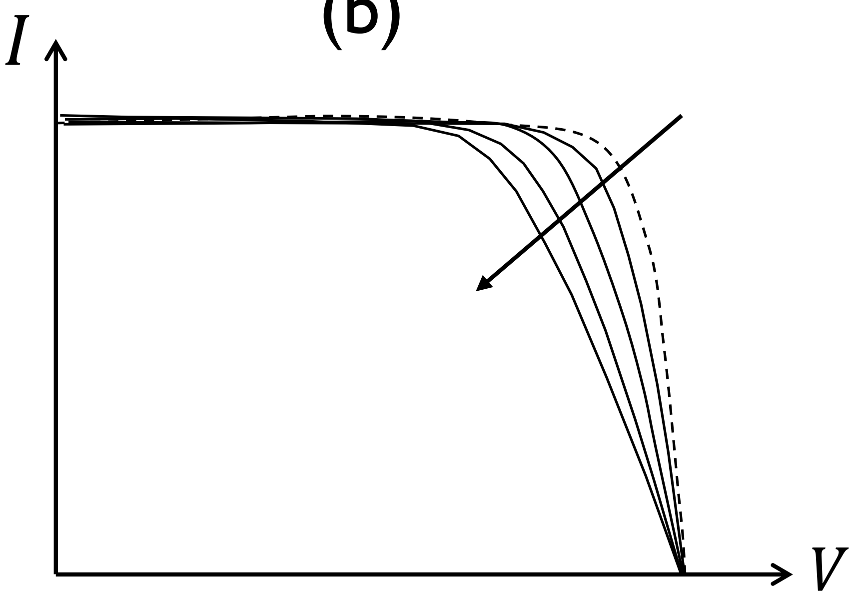
Degradations occur in parallel

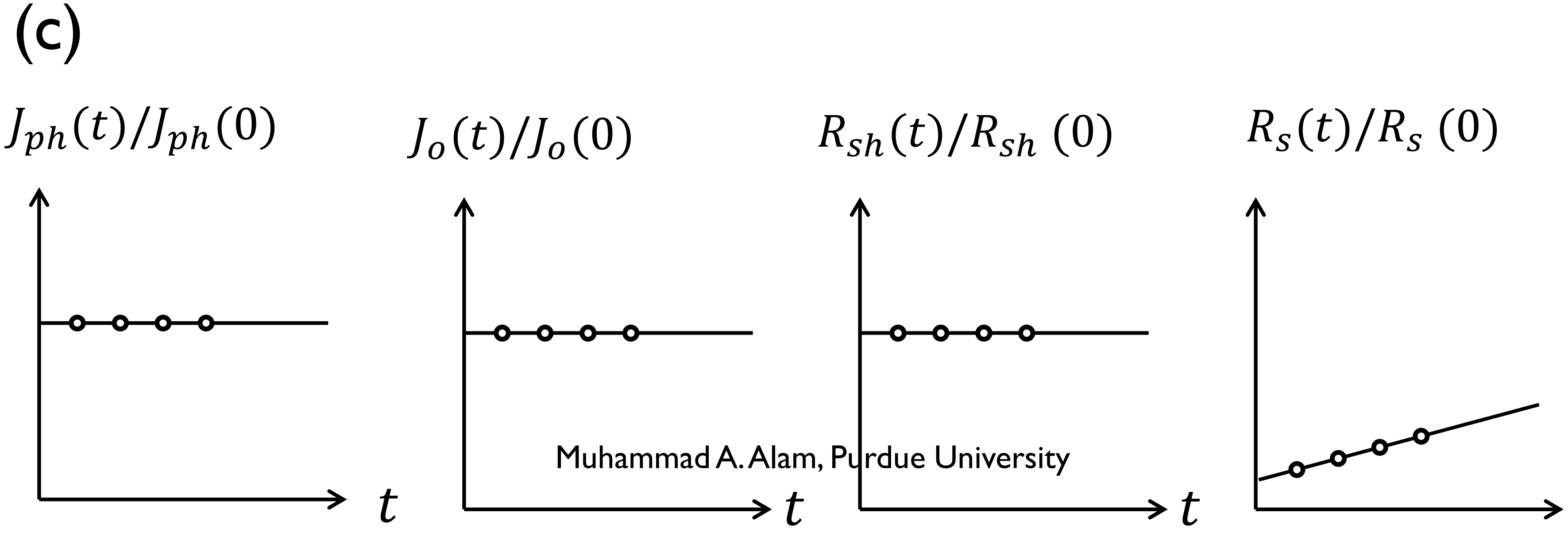


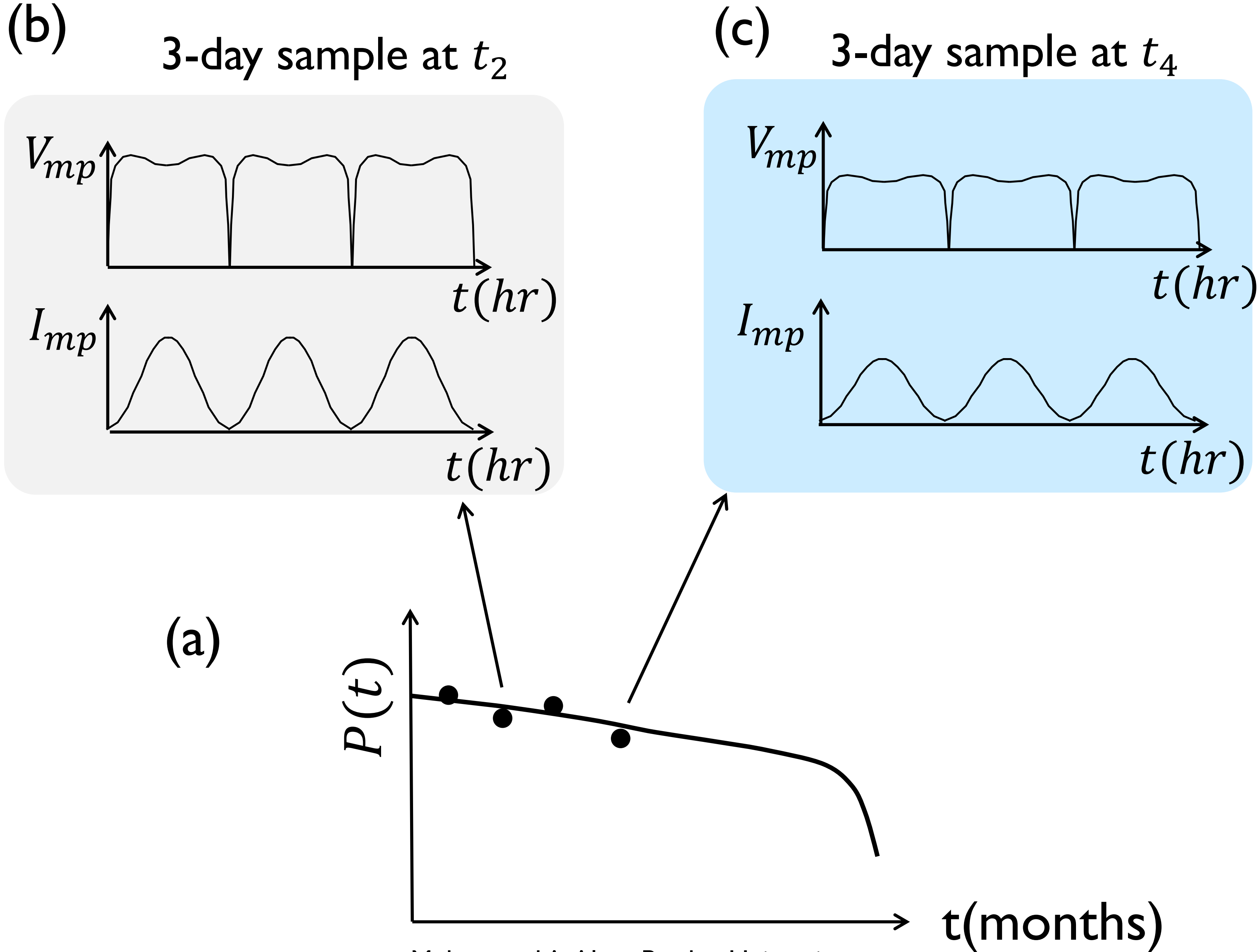




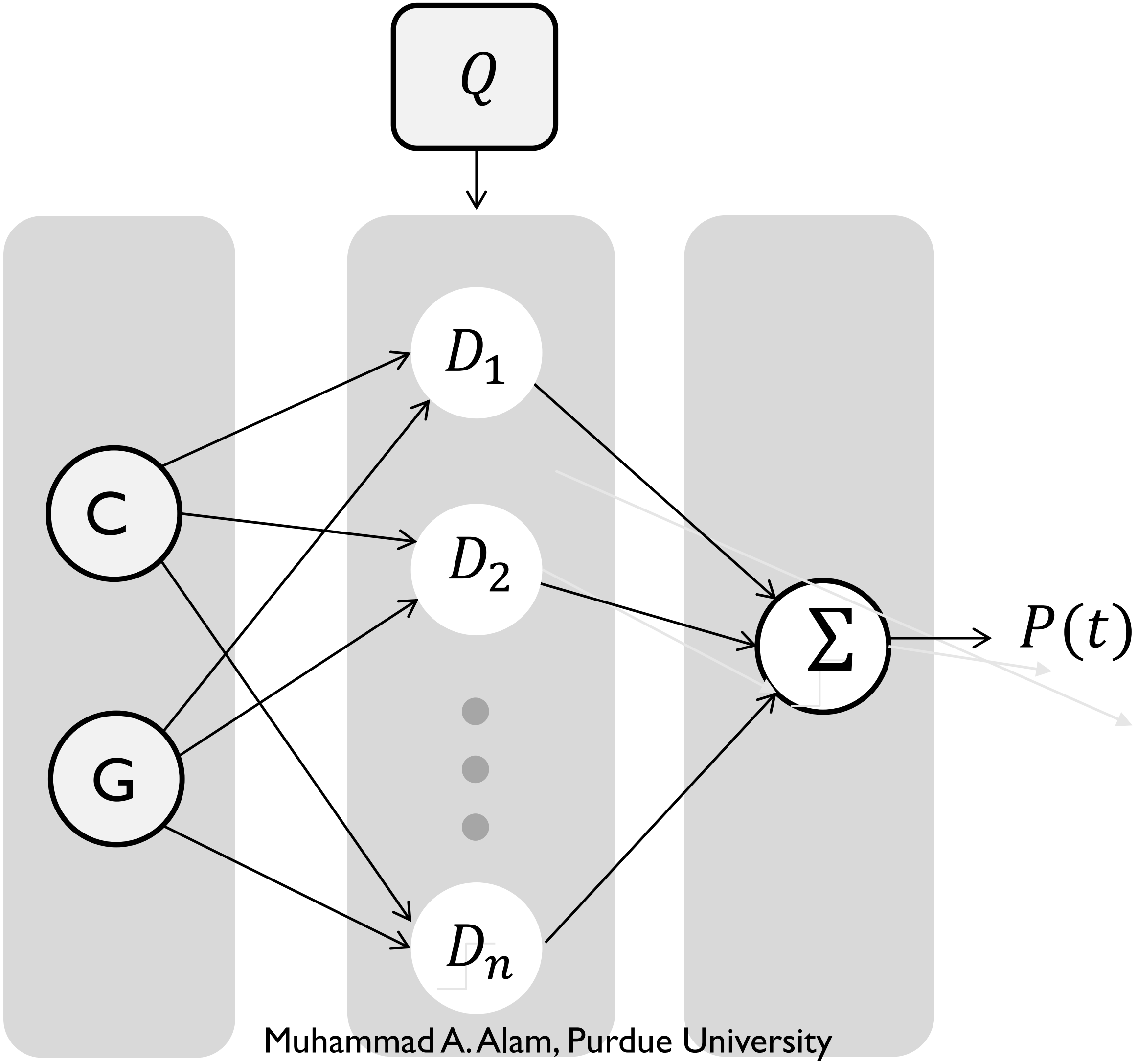

(a)
$P(t)$
P_{crit}
t_0 time


(b)
I
V

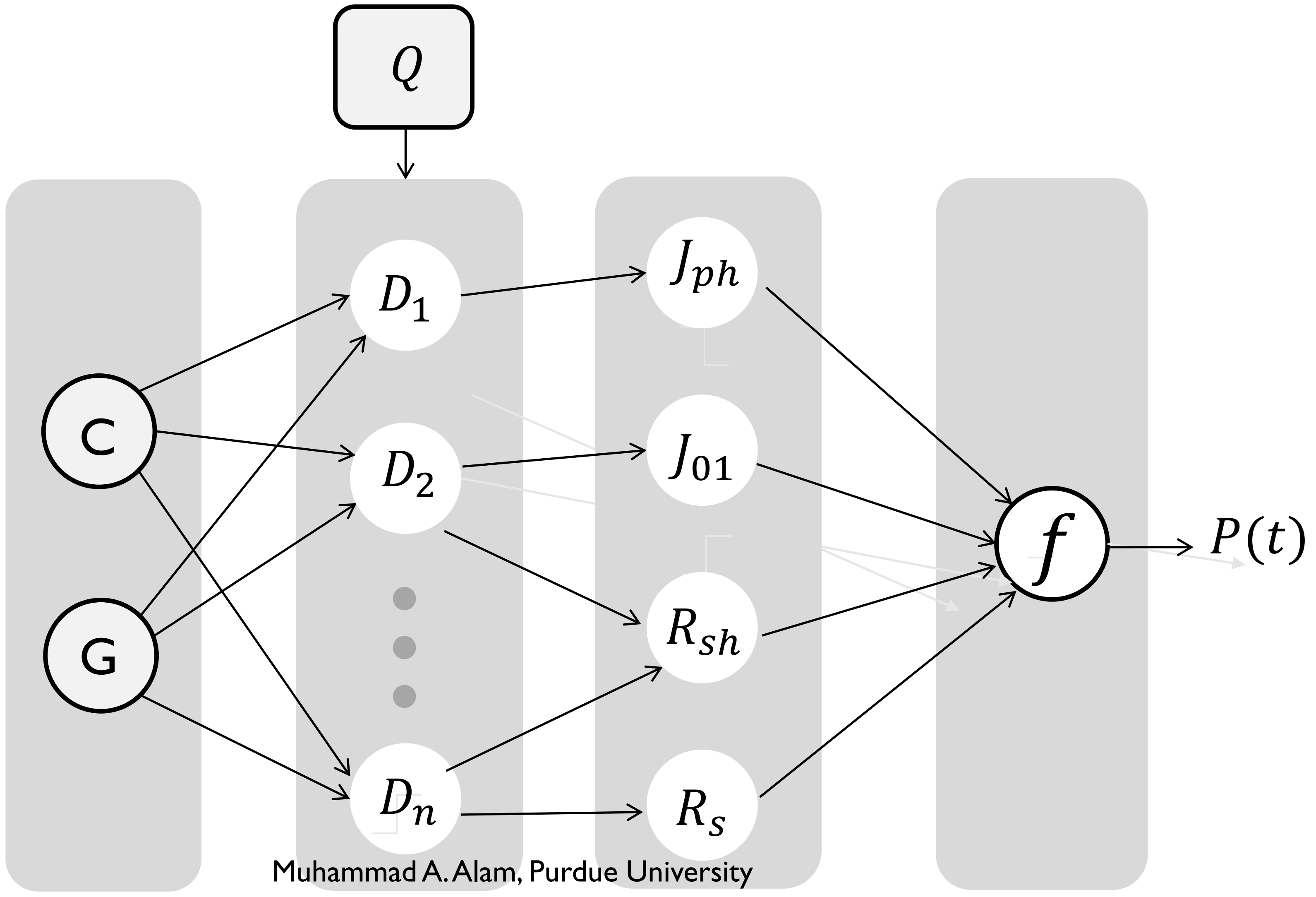

(c)
$J_{ph}(t)/J_{ph}(0)$
$J_o(t)/J_o(0)$
$R_{sh}(t)/R_{sh}\ (0)$
$R_s(t)/R_s\ (0)$
t
t
t
t


(b)
3-day sample at t_2
V_{mp}
$t(hr)$
I_{mp}
$t(hr)$
(c)
3-day sample at t_4
V_{mp}
$t(hr)$
I_{mp}
$t(hr)$
(a)
$P(t)$
t(months)

Additive power degradation model

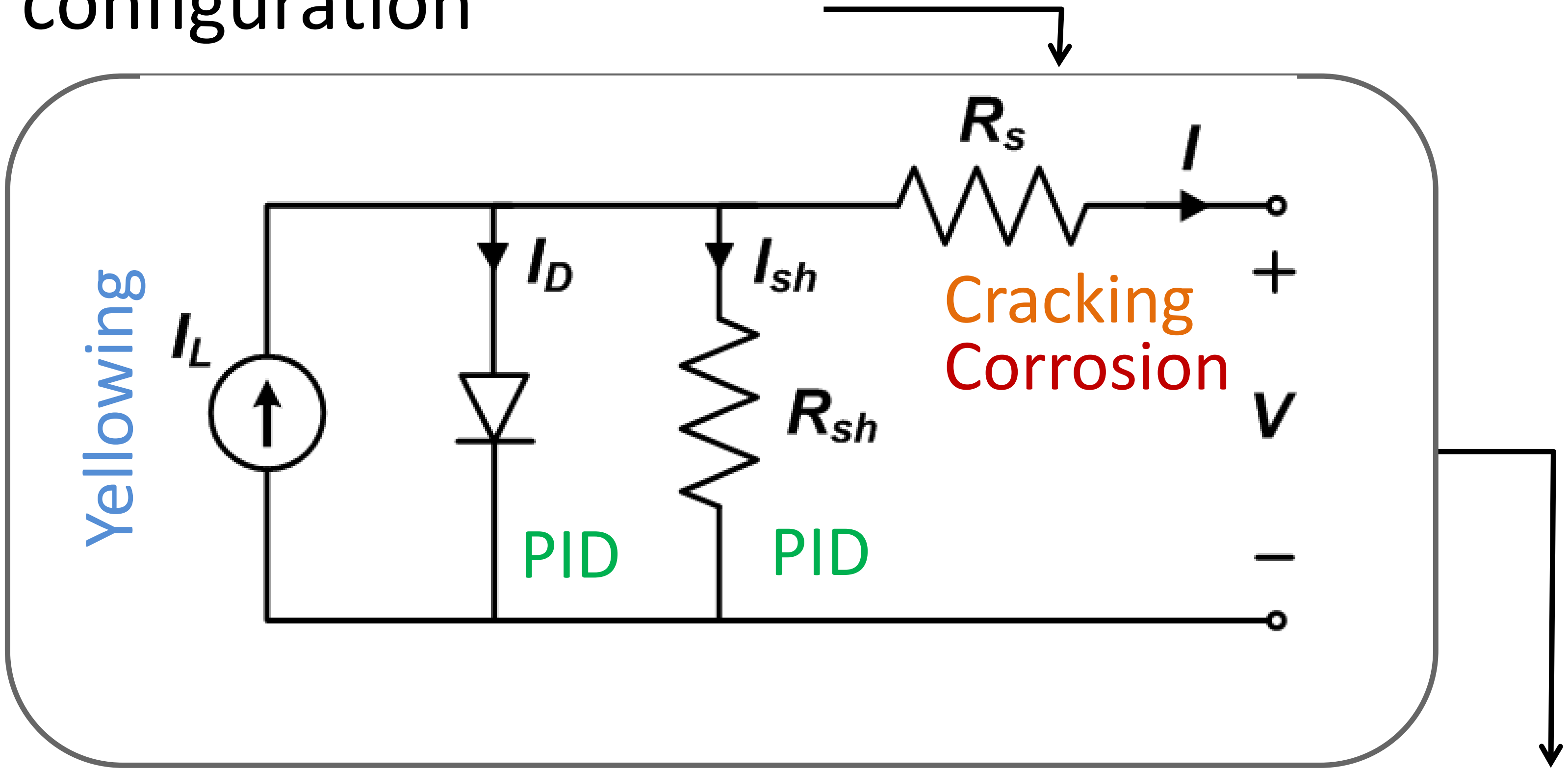


Compact model-based model



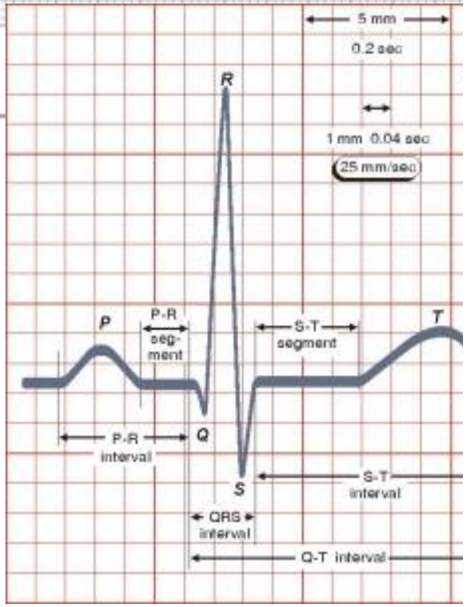
Compact model based approach

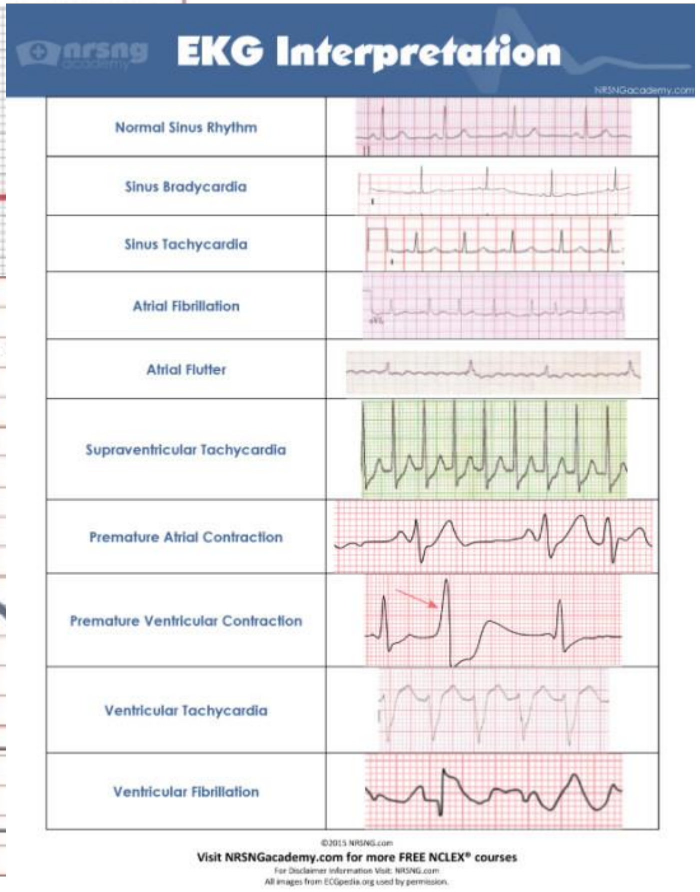
Weather & cell/module/farm configuration



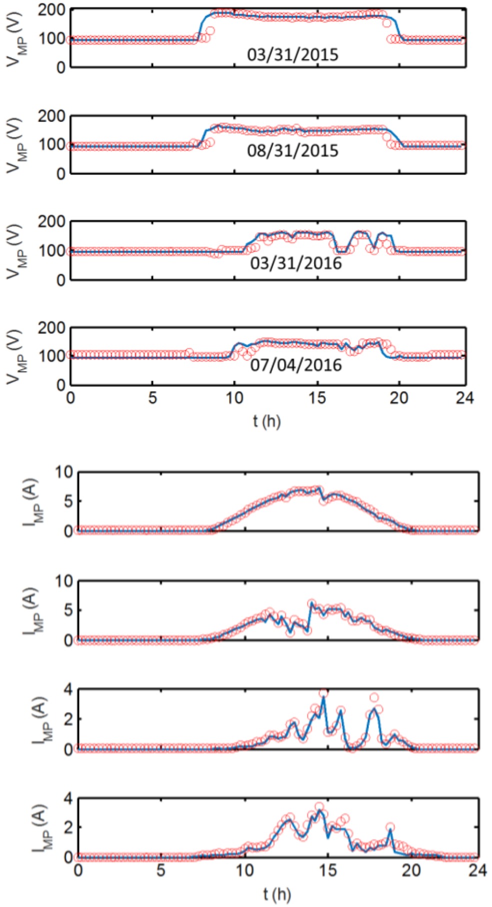
Time-dependent power output

Approach: An EKG for solar

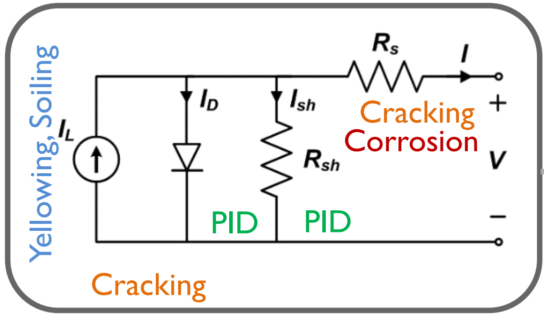




Approach: PV 'Heartbeat' Interpreted



Temperature, RH, technology, module configuration



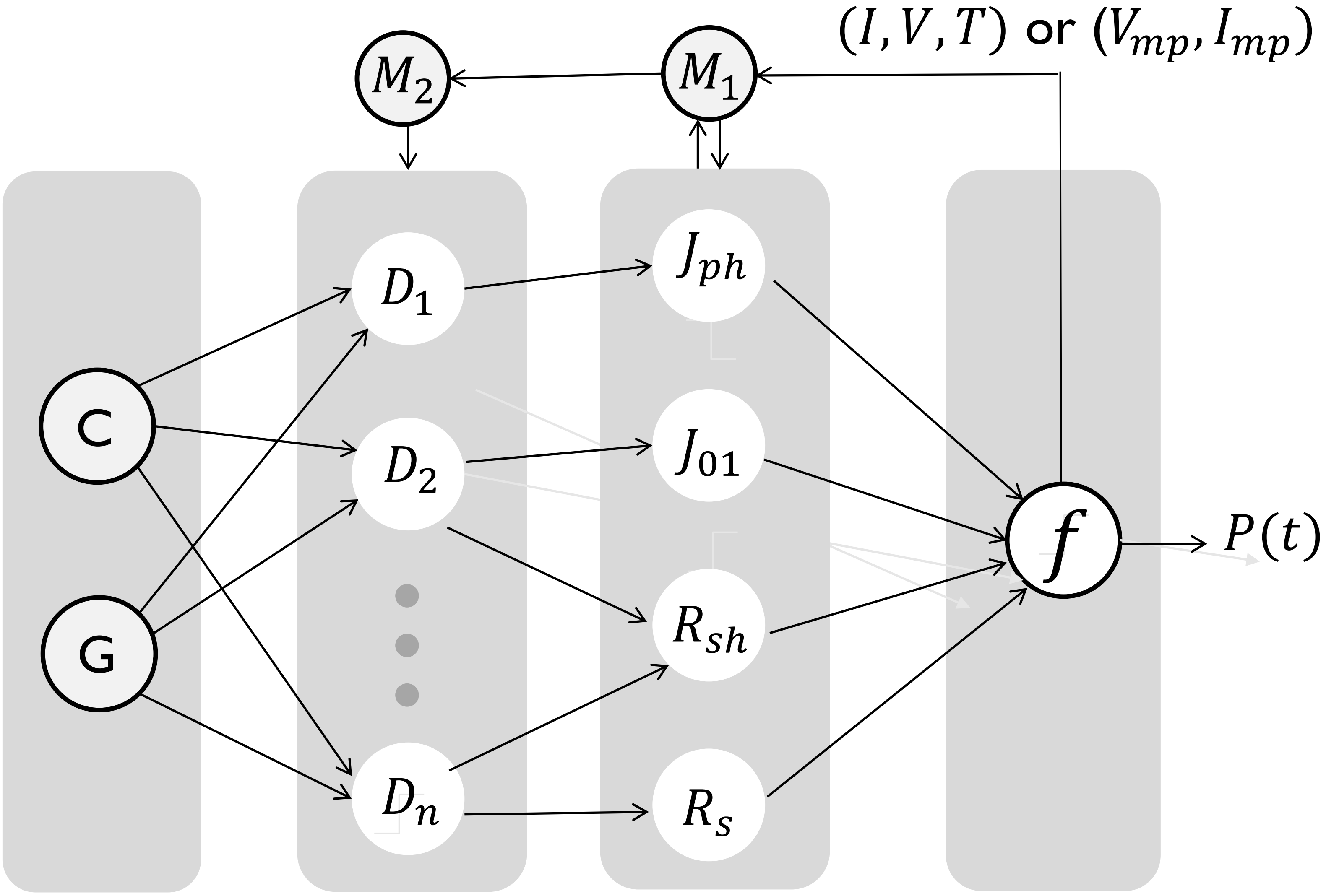
- 24 Solar Panels installed at Purdue

- Archived field data every 15 mins for 3 years

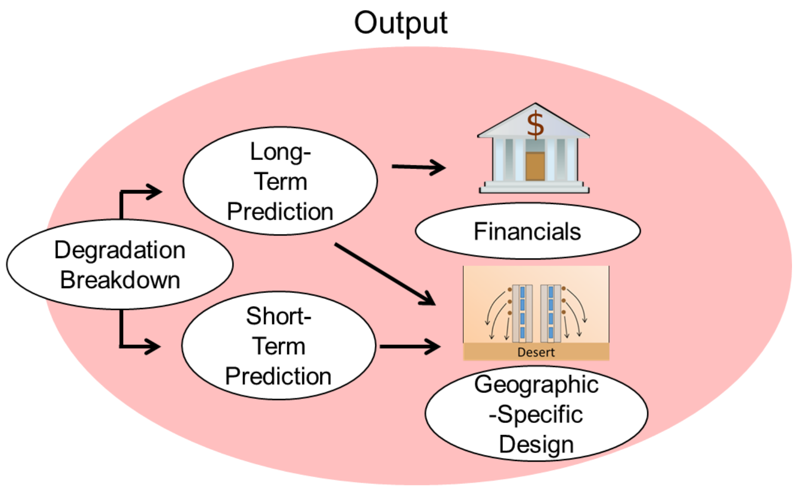
Time-dependent power output

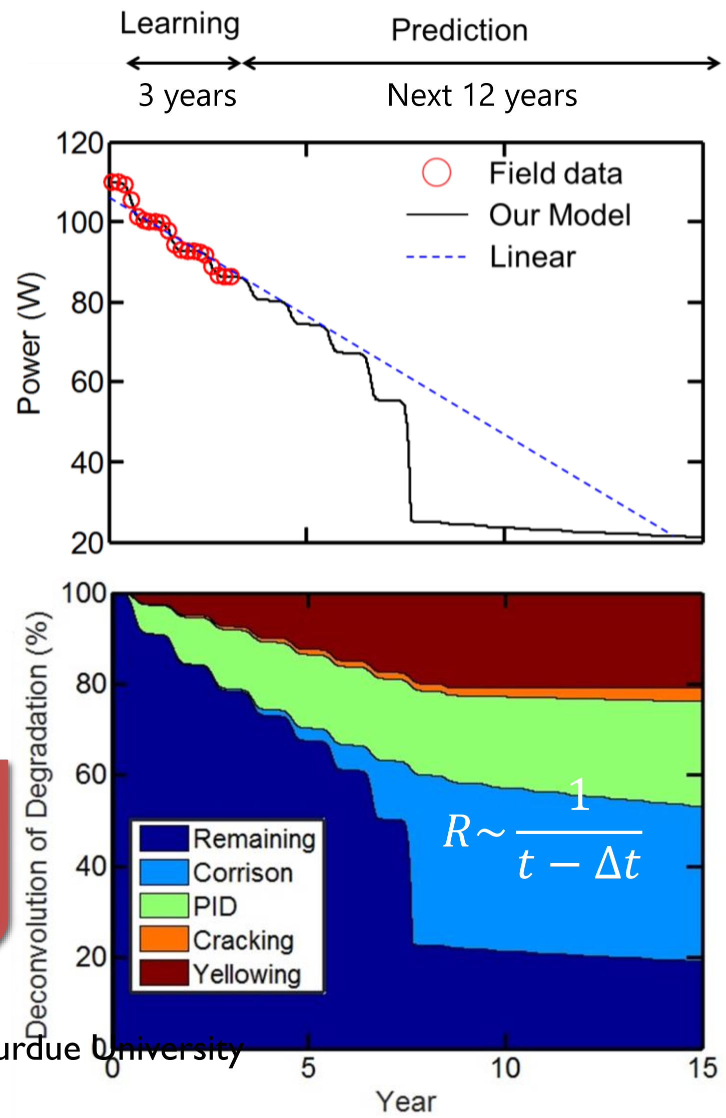
Combining environmental and module data achieves excellent match with experiment

Physics-based machine learning approach



Approach: Degradation deconvolution and lifetime prediction





Our framework allows inverse modelling of ***3-year*** simulated field data, and predict the energy yield for total ***15 years***.