

# RAPID LIFE-CYCLE IMPACT SCREENING FOR DECISION-SUPPORT AT EARLY STAGE CHEMICAL DESIGN

RUNSHENG SONG, ARTURO A. KELLER AND SANGWON SUH\*

BREN SCHOOL OF ENVIRONMENTAL SCIENCE AND MANAGEMENT,  
UC SANTA BARBARA,  
USA

*Environmental Science & Technology;  
In Revision*

# About me



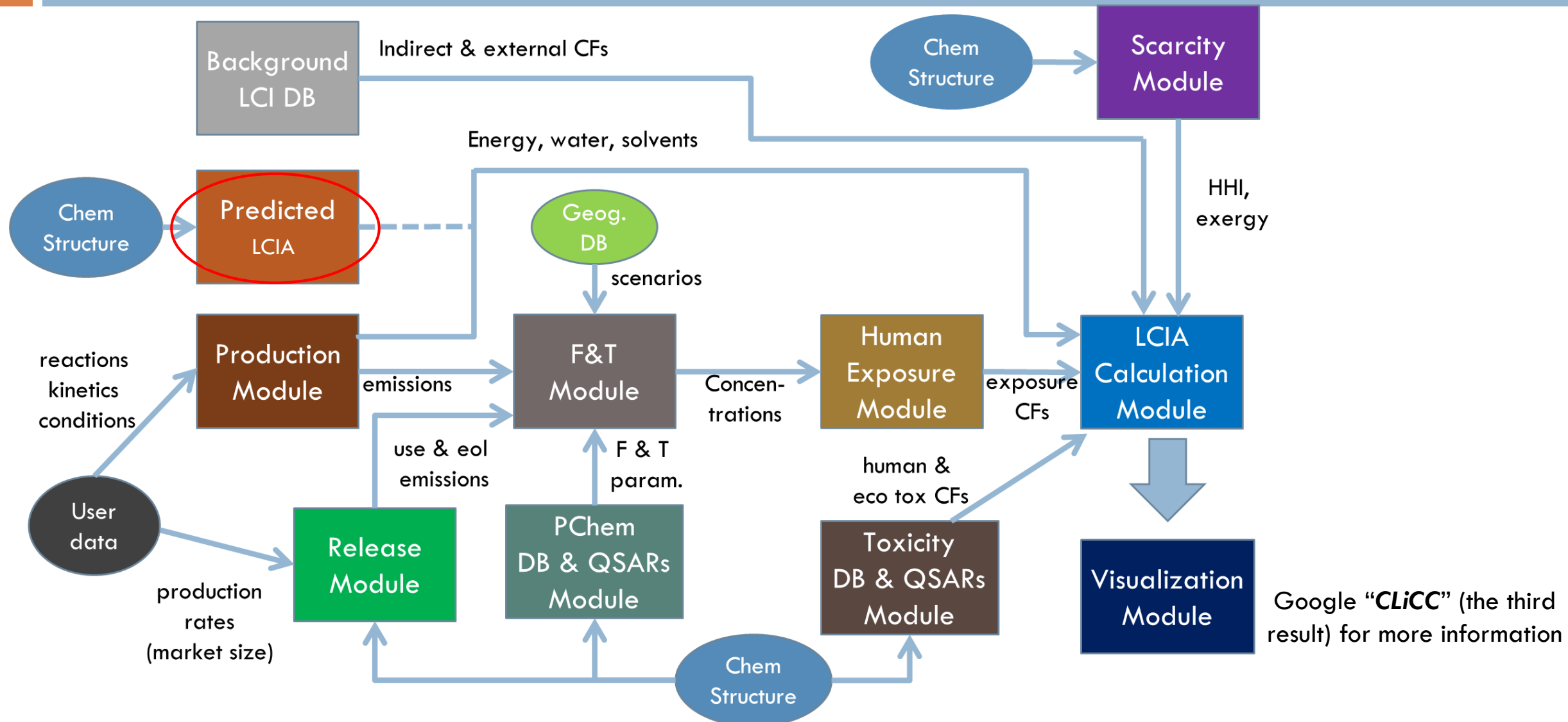
- Third year PhD Student
  - Prof. *Sangwon Suh* and Prof. *Arturo Keller*
- Life-Cycle Assessment for Chemicals; Life-Cycle Inventory Database; Machine Learning; Chemical Toxicity Prediction.
- Chemical Life-Cycle Collaborative (CLiCC) with US EPA.



# Background – The CLiCC Project

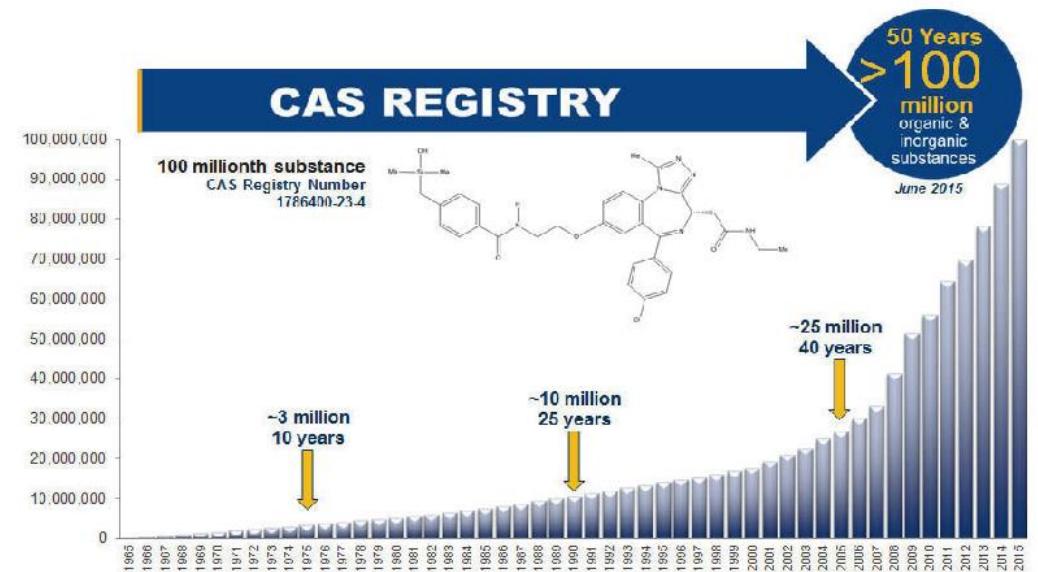


CHEMICAL LIFE CYCLE  
COLLABORATIVE



# Background – Predictive LCIA

- There are more than **100 million** chemicals in CAS database;
  - ▣ **15,000** new chemicals are being added everyday.
- In many cases, we don't have the necessary data to build LCI for chemicals;
- Alternative path to estimate LCA indicators.
  - ▣ e.g., CED, GWP and Eco-indicator;
  - ▣ Estimates the indicators with molecular structure information using machine learning models.



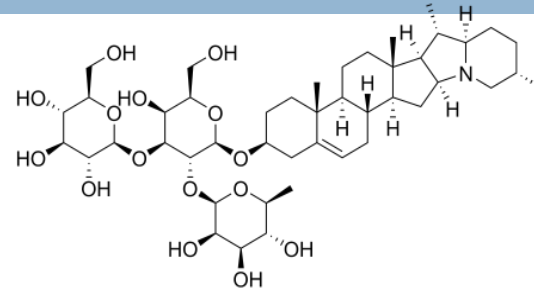
No data received

[Details](#)

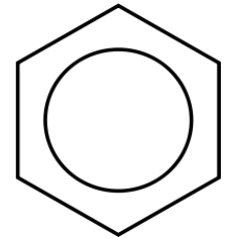
Reload

# Background – Predictive LCIA

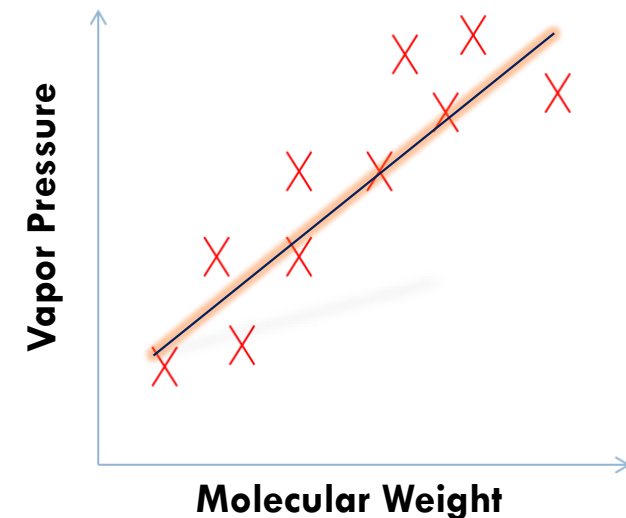
- Chemical structure is correlated with chemical properties and impacts;
- Linear regression model has been widely used to approximate chemical impact;
  - ▣ The predictive power is restricted.
- Nonlinear model shows better predictive power
  - ▣ Artificial Neural Networks (ANNs) outcompetes linear regression model in estimating life-cycle indicators for chemicals.



*Might consume more energy*

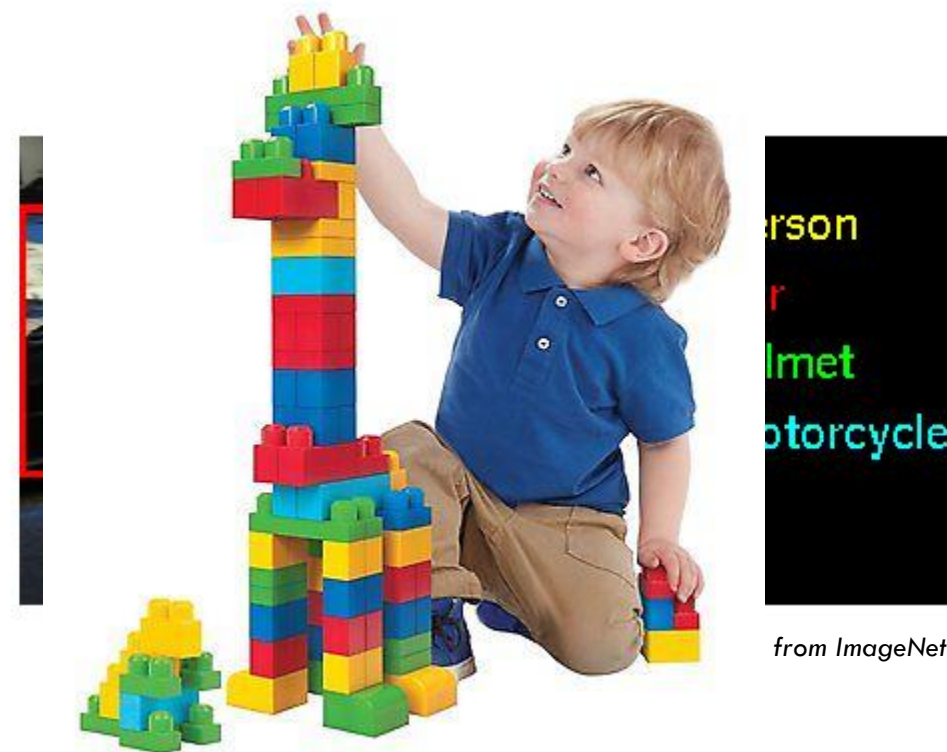
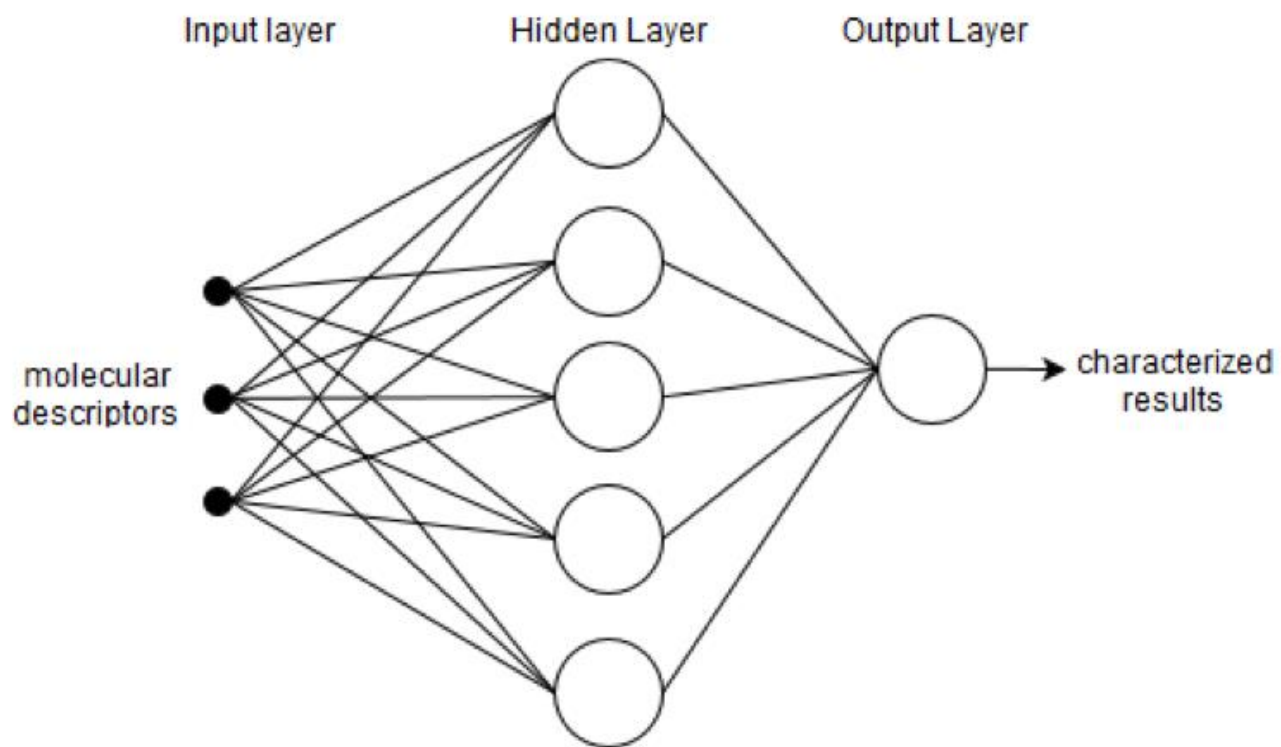


*...than this*



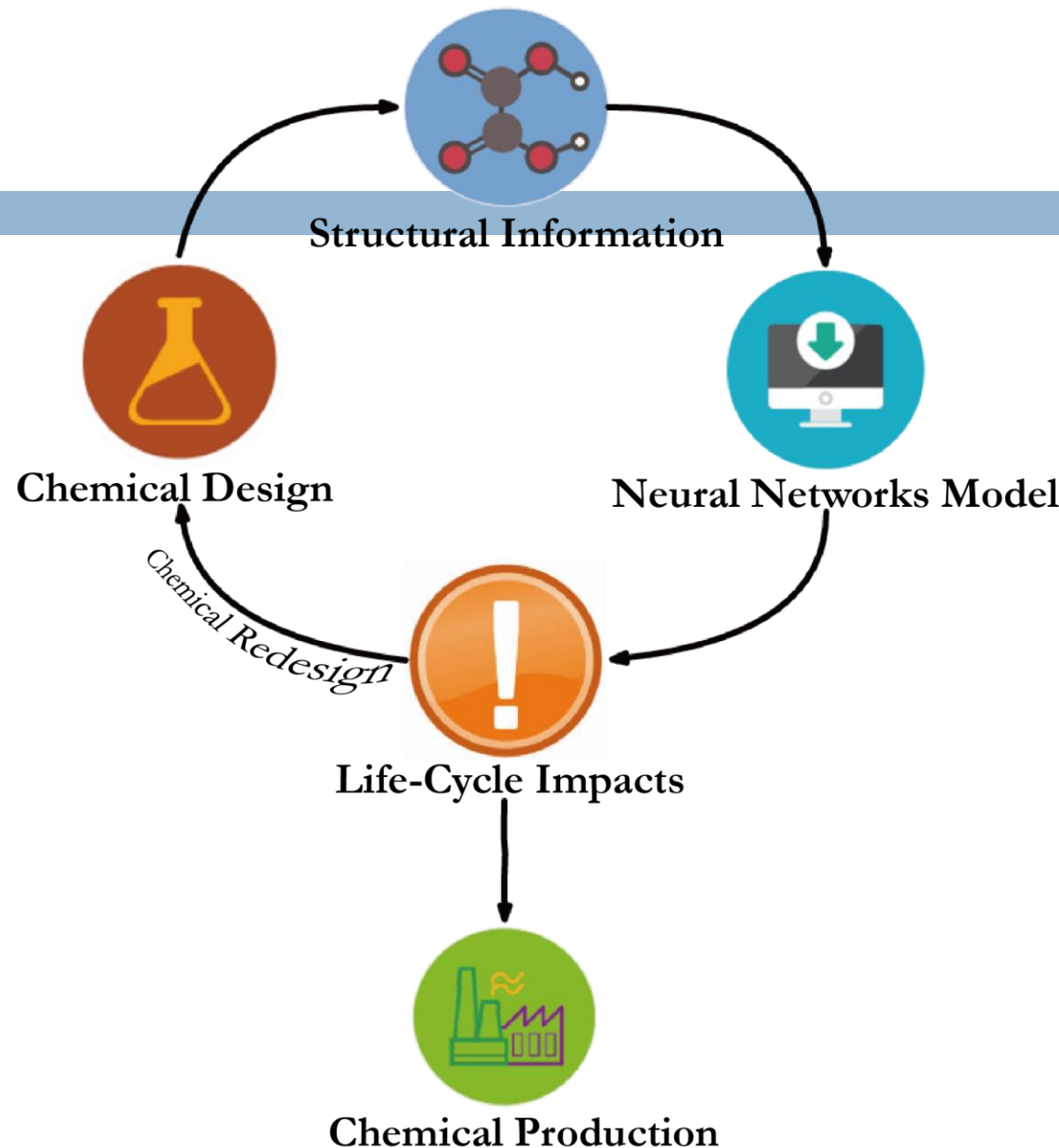
# Background – Deep ANNs

- ANNs model becomes very popular because of the concept “*deep learning*”



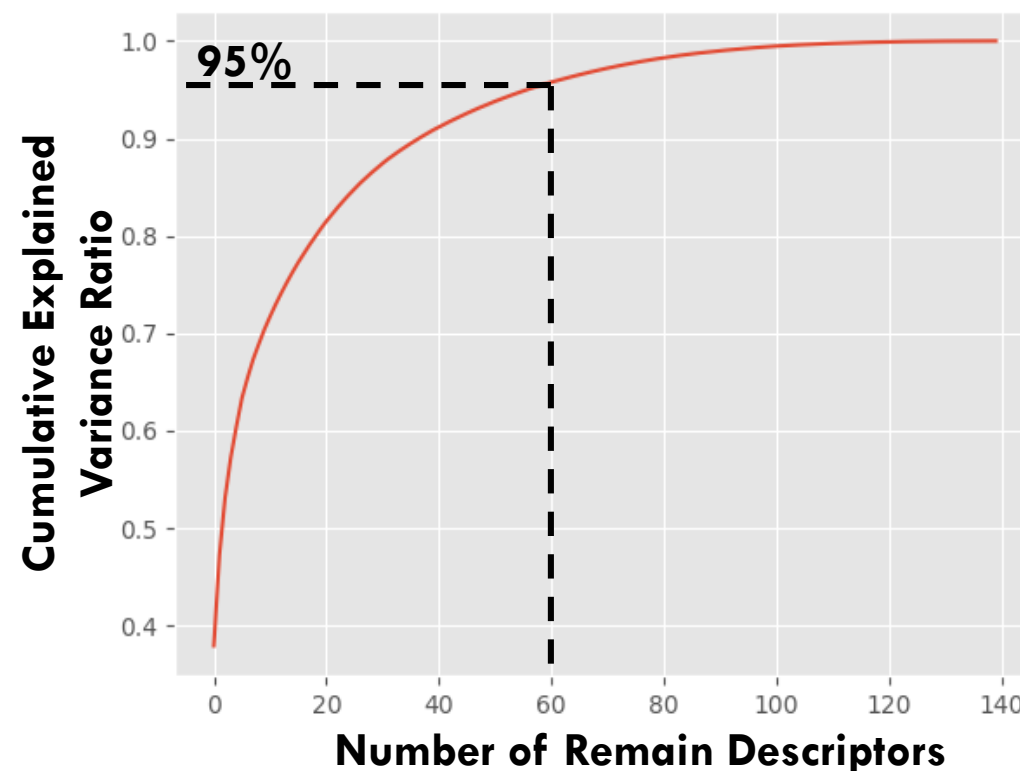
# Highlights of This Study

- Estimate the life-cycle impact indicators for chemical
  - ▣ Use deep Neural Networks model;
  - ▣ Use high dimensional molecular structure descriptors;
  - ▣ Model structures were tuned;
  - ▣ Model Applicable Domains (AD) were measured;
  - ▣ What we learnt from this study.



# Method – Data

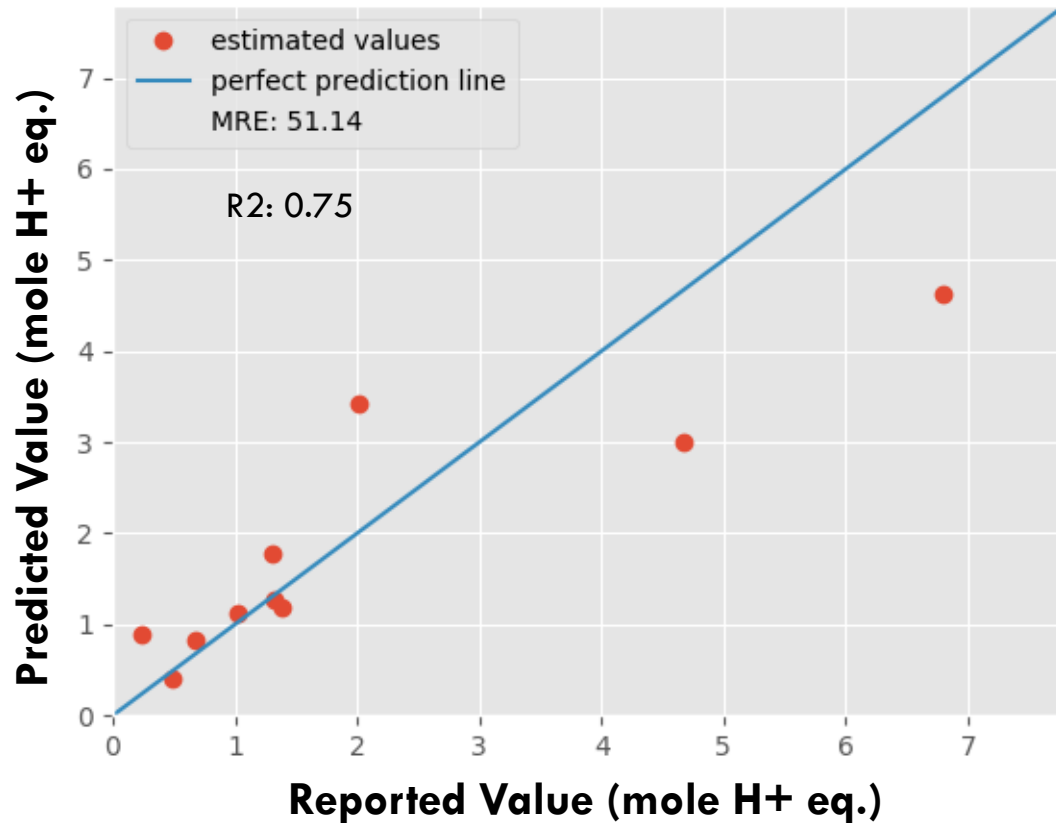
- 166 chemical LCI data were collected from Ecoinvent v3.01
  - 10 chemicals were used as testing set
  - 10% of the rest 156 chemicals were validation set
- About 4,000 molecular descriptors were generated by software Dragon 7.
- Principle component analysis was used to reduce the dimension of the descriptors.



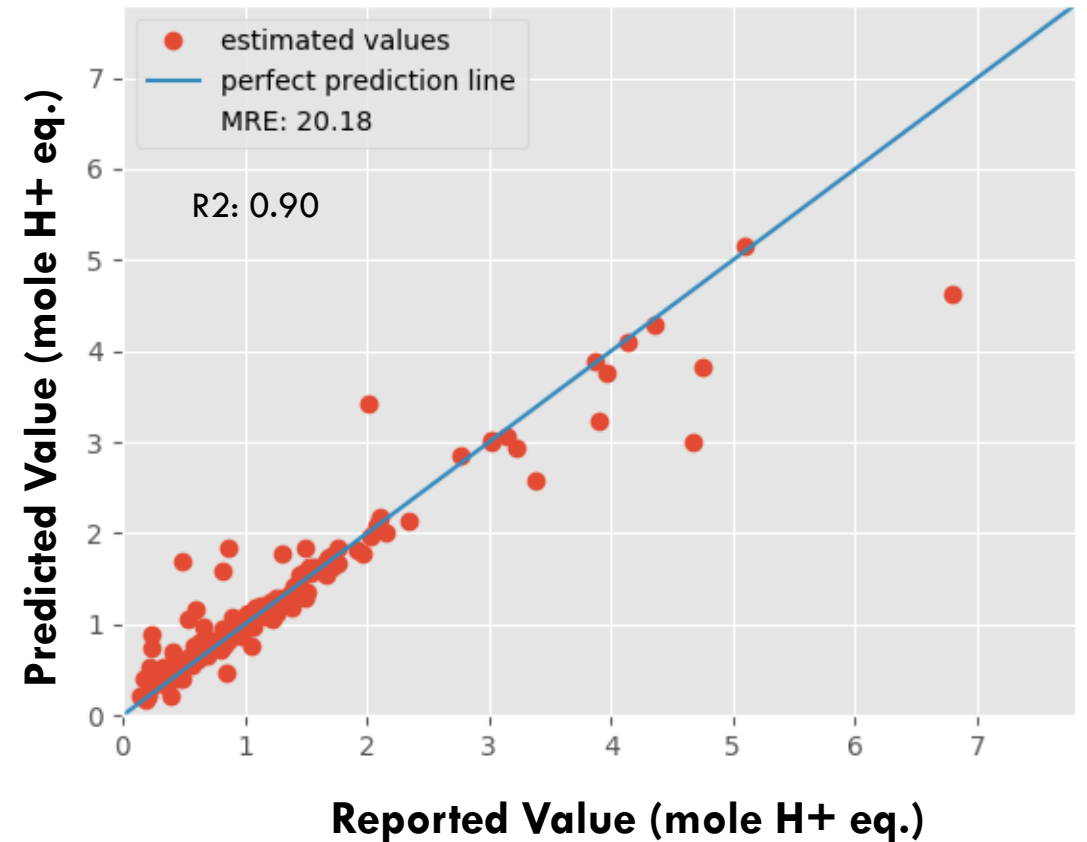


# Results – Acidification Model

## Only Test Set

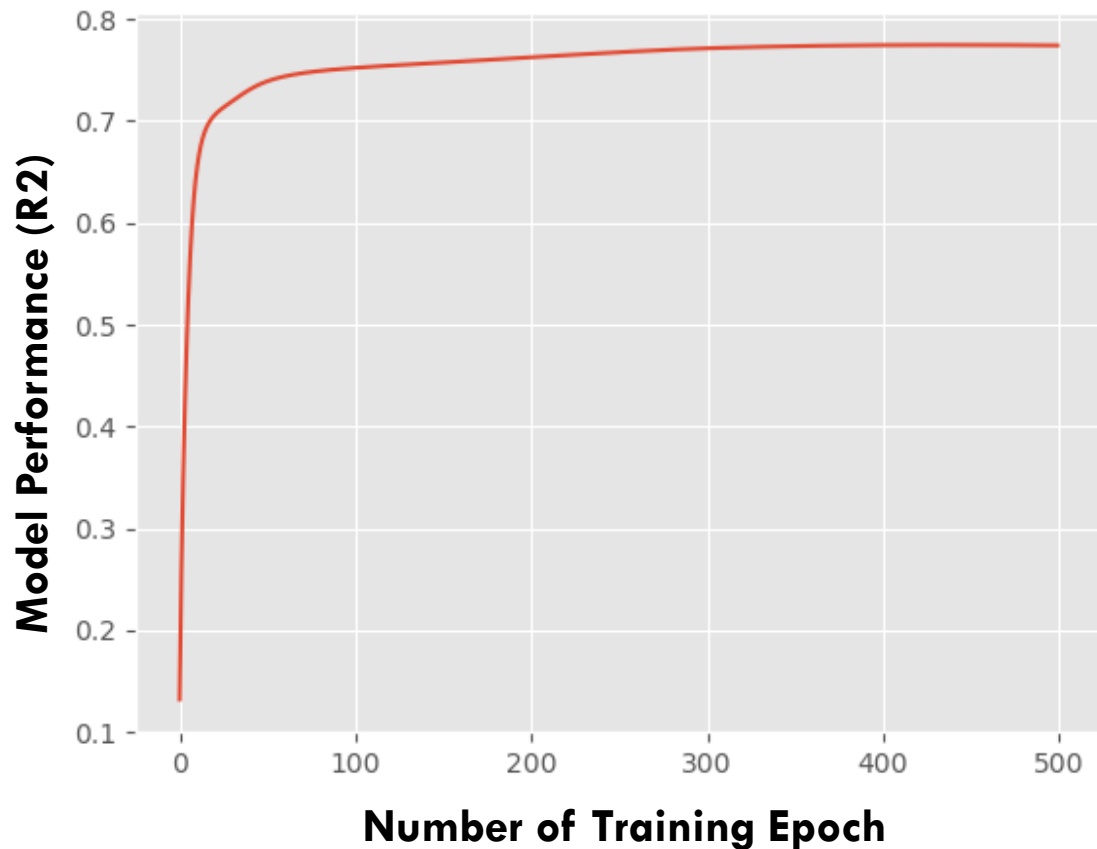


## All Chemical We have...

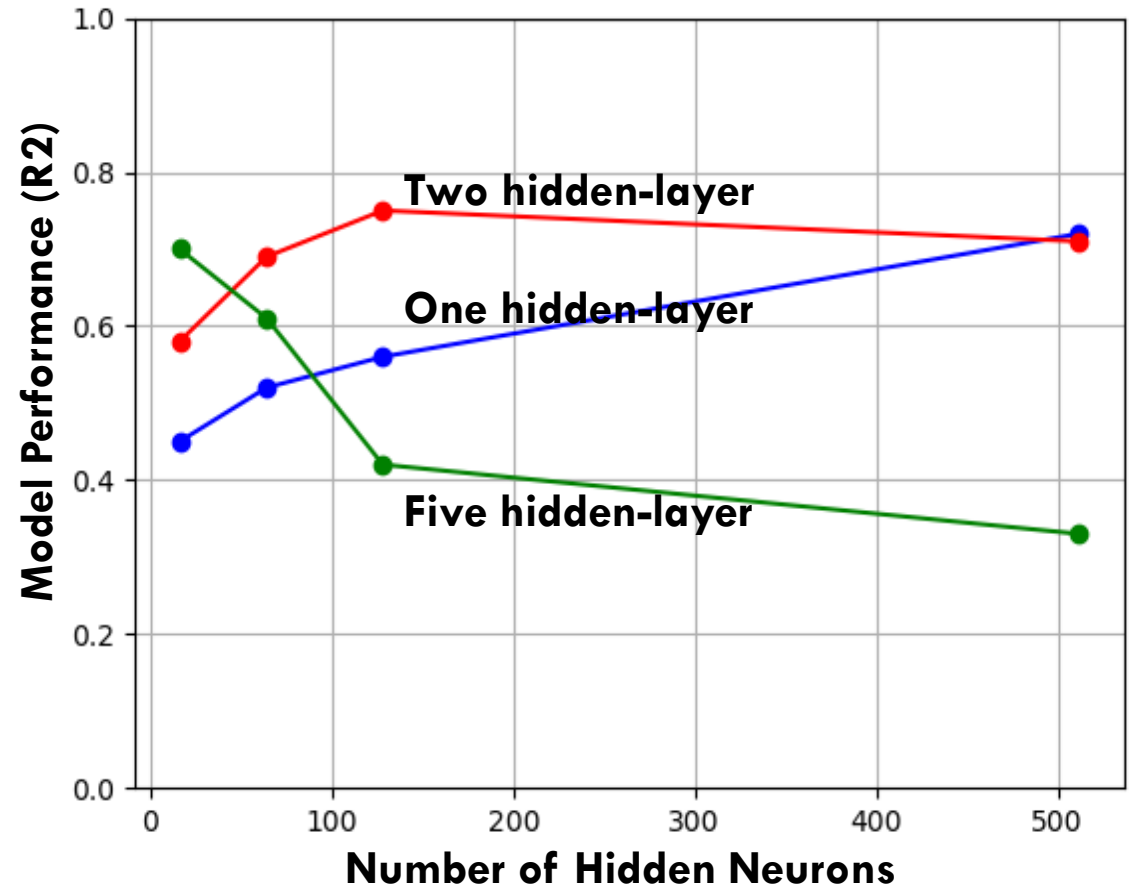


# Results – Model Training

### Training Process of Acidification Model

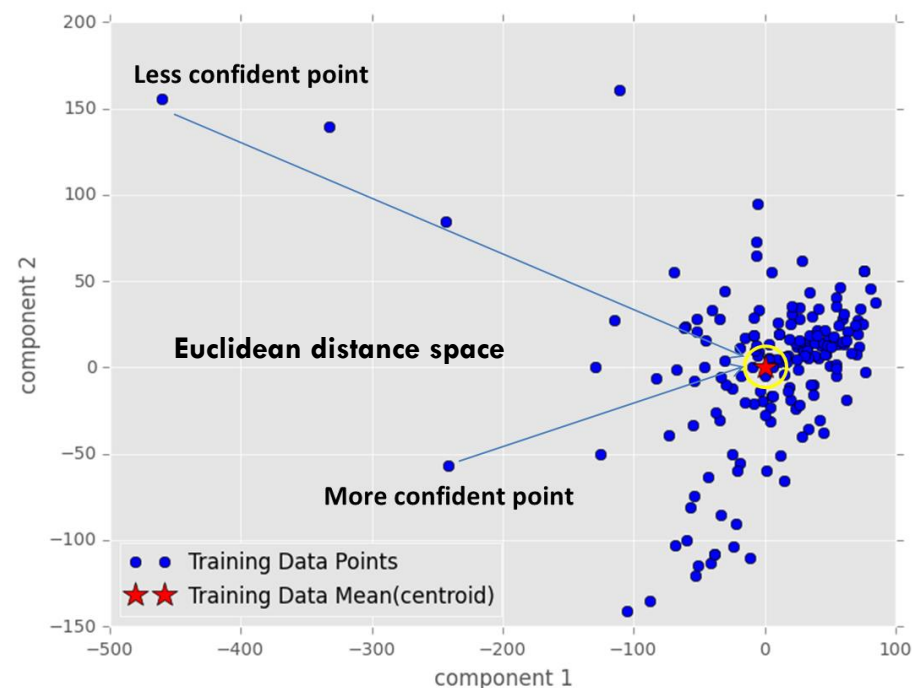


### Cross-Validation for Acidification Model



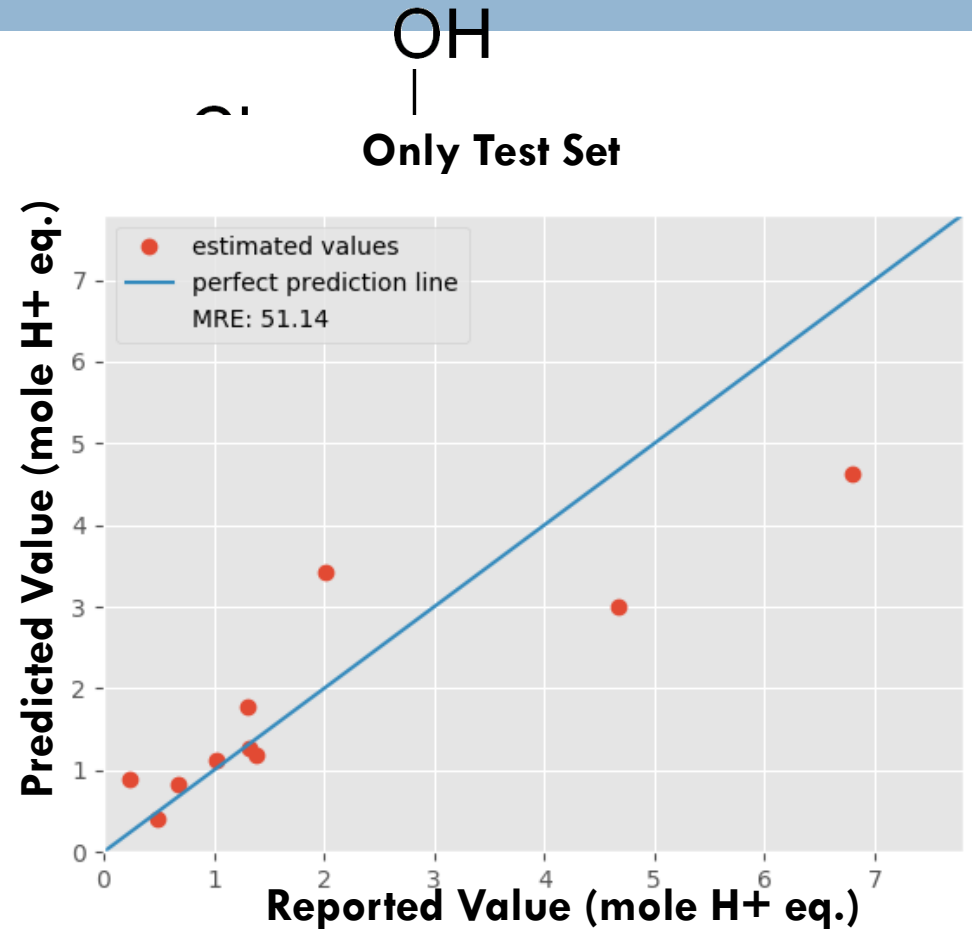
# Model Applicable Domain

- Query chemicals that have higher structural similarity with the training data are likely to have higher prediction accuracy.
- Accuracy could be measured depending on if this chemical falls into the applicable domain.
  - ▣ MRE of test chemical inside AD: 40%
  - ▣ MRE of test chemical outside AD: 85%



# Model Demo on Test Data

- 2,4-Dichlorophenol
  - ▣ TRACI, Acidification: 1.32 (moles of H<sup>+</sup>-Eq);
  - ▣ Our model estimates: 1.27 (moles of H<sup>+</sup>-Eq);
  - ▣ Uncertainty according to AD: Low
  
- Hexafluoroethane
  - ▣ TRACI, Acidification: 6.8 (moles of H<sup>+</sup>-Eq);
  - ▣ Our model estimates: 4.6 (moles of H<sup>+</sup>-Eq);
  - ▣ Uncertainty according to AD: High

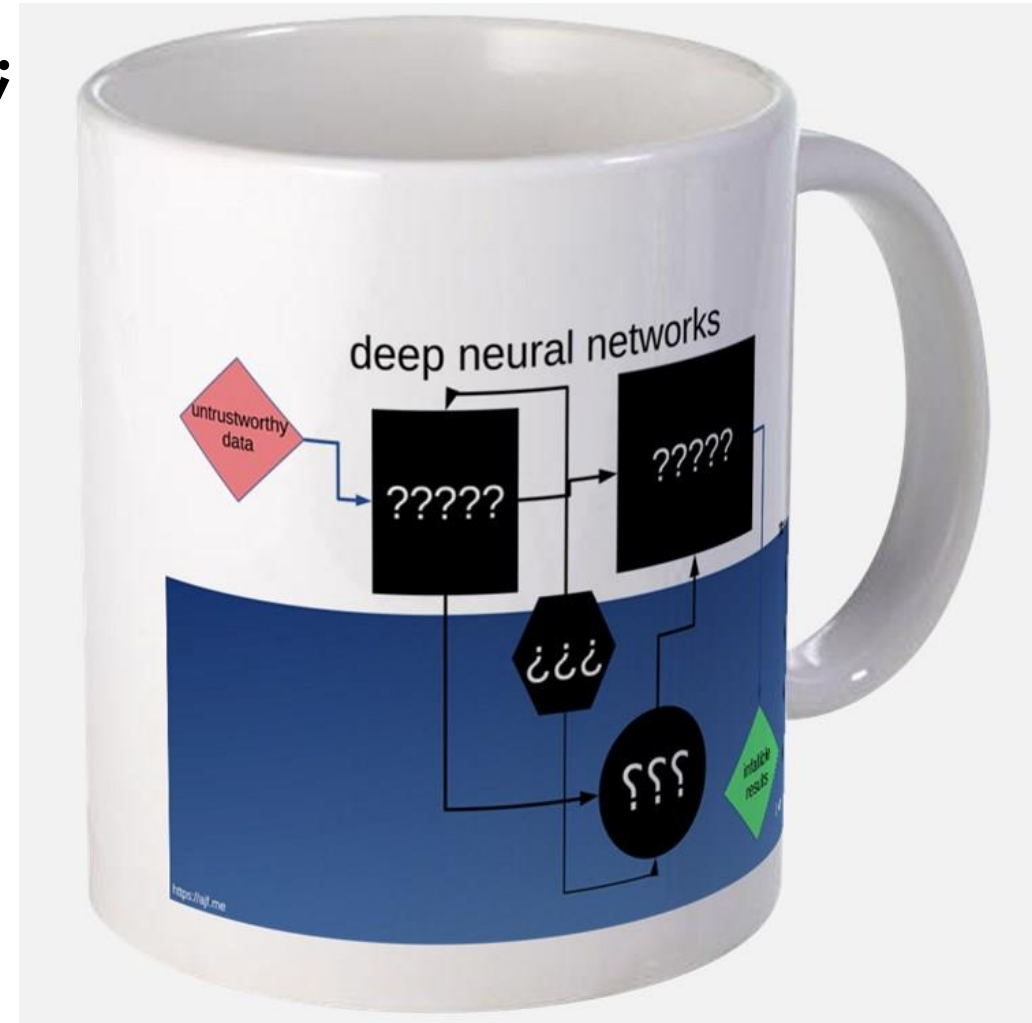


# Conclusion

- We are able to predict three mid-point indicators (CED, Acidification, GWP) and three end-points indicators (EI99, Human health, Ecosystem quality);
- The cross-validated models show good predictive power on testing data ( $R^2 > 0.7$ );
- Model applicable domain measurement can indicate the uncertainty of the prediction;
- The end-point indicators require higher complexity of the model.

# Outlooks

- More training data will always be beneficial;
- It's hard to tell the contribution of each input descriptors;
- This field is developing very fast.

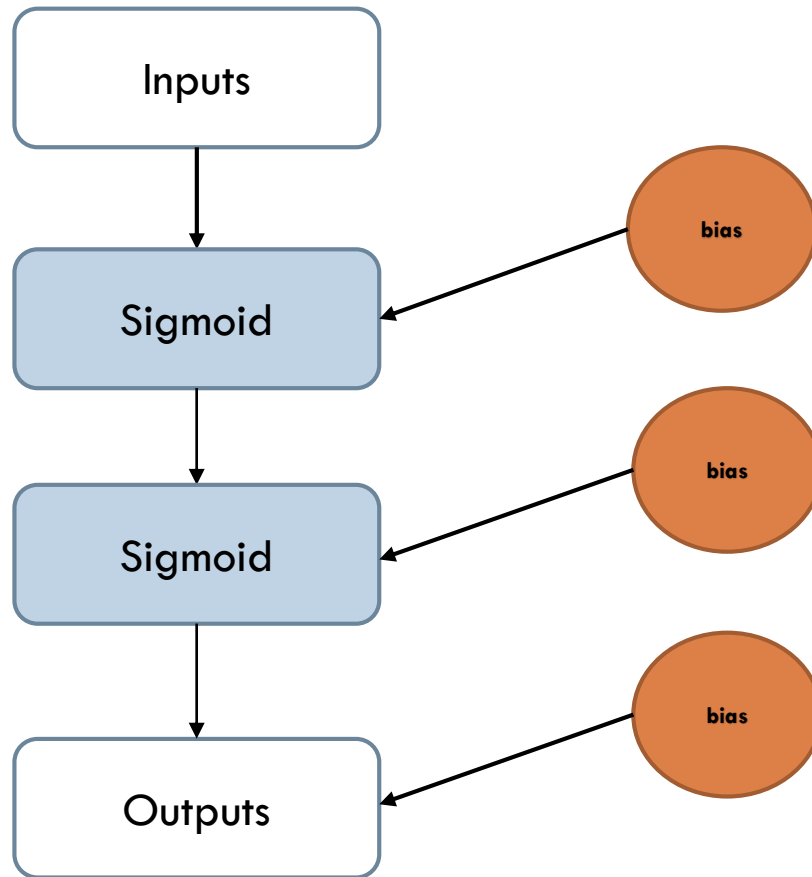


# Thank you



BREN SCHOOL OF  
ENVIRONMENTAL SCIENCE  
& MANAGEMENT  
University of California, Santa Barbara







### Cross-Validation for Acidification Model

