

DATA SCIENCE CONSULTING SERVICE

As part of **Purdue Integrative Data Science Initiative (IDSI)**, **Data Science Consulting Service** will provide hands-on consulting support for **data analysis** and **business analytics** in all areas to overcome data science challenges arising in research, education, and business and organization management. Our consultants have advanced degrees and years of experience in **deep machine learning**, **data mining**, **big data analysis**, **artificial intelligence**, **business analytics** and **computational statistics**.

Guang Lin, Director, Data Science Consulting Service,
Associate Professor, Departments of Mathematics/Mechanical Engineering/ Statistics (Courtesy)

Supported by IDSI & College of Science

Special thanks go to: Patrick Wolfe, Sunil K. Prabhakar, Pankaj Sharma, Daniel E. Hirleman, Bruce Craig, George McCabe, Greg Buzzard, Hao Zhang, Larry S. Sommers, Preston M. Smith, Adrienne N. Thompson, Holly G. Graef, and Pamela G. Burroff-Murr.



SERVICES AND EXPERTISE

- **Data Science Consulting Service (DSCS)** faculty, staff and consultants aim to provide individualized consulting service to Purdue faculty, research staffs, students and industrial/business clients, offering hands-on consulting service on resolving their data science related questions. The consulting fee for Purdue faculty, staffs and students is covered by **Integrative Data Science Initiative (ISDI)** and **Purdue College of Science** for the year of 2019. We welcome Purdue faculty, research staffs, students, and industry clients to contact us directly to discuss their research with us.
- Our faculty, staff, and consultants are available, **by appointment**, to help you with data science related issues.



EXPERTISE

- We provide **hands-on consulting service** on data management, business analytics, high-performance data processing and the use of advanced data-science, machine learning, big data analysis, and artificial intelligence tools and methodologies to analyze client's datasets. In particular, our expertise lies in:
- **Data and information management.** DSCS faculty, staff and consultants have experience in using advanced database and data processing tools to manage big, and unstructured data for analysis using a variety of scripting languages and tools.
- **Advanced methodology for data science.** DSCS faculty, staff and consultants have expertise in methodological aspects of data science including statistical data analysis, machine learning, artificial intelligence, uncertainty quantification, and sensitivity analyses.
- **Data exploration.** DSCS faculty, staff and consultants have experience in data visualization, interpretation, and hypothesis-generating research.
- **High-performance data processing.** DSCS faculty, staff and consultants have expertise in optimization of code for data processing in CPU and GPU environment.
- **Business analytics and business intelligence.** DSCS faculty, staff and consultants have expertise in using advanced statistical analysis, data mining, machine learning and artificial intelligence tools to explore the client's data in support of data-driven decision-making.
- **Grant proposal and manuscript preparation.** DSCS faculty, staff and consultants will work with clients to write sections related to data science of grant proposals or manuscripts as co-PI/senior investigators or co-authors to make grant application or manuscript submission more competitive.

CHOOSE THE RIGHT SERVICE

- If you are a [Purdue researcher](#) who seeking statistical assistance on a research project, the [Statistical Consulting Service](#) is the right service for you.
- If you are an [industrial/business client](#) looking for business analytics and business intelligence solutions, the Data Science Consulting Service (DSCS) is the right service for you. We will employ advanced data-science tools to explore your dataset in support of data-driven decision-making.
- If you are an [industrial/business client](#) or [Purdue researcher](#) and would like data science experts to analyze your datasets directly and provide you with the customized technical report, DSCS is the right service for you.
- If you are an [industrial/business client](#) or [Purdue researcher](#) and would like data science experts to develop some predictive tools using the state-of-the-art deep learning software such as TensorFlow or PyTorch, DSCS is the right service for you.
- If you are an [industrial/business client](#) or [Purdue researcher](#) and do not have the computational power equipment and data management tools, DSCS is the right service for you. We have the access to the powerful GPU-clusters and state-of-the-art data management and processing tools to manage and analyze your big datasets on a high-performance parallel data processing platform.



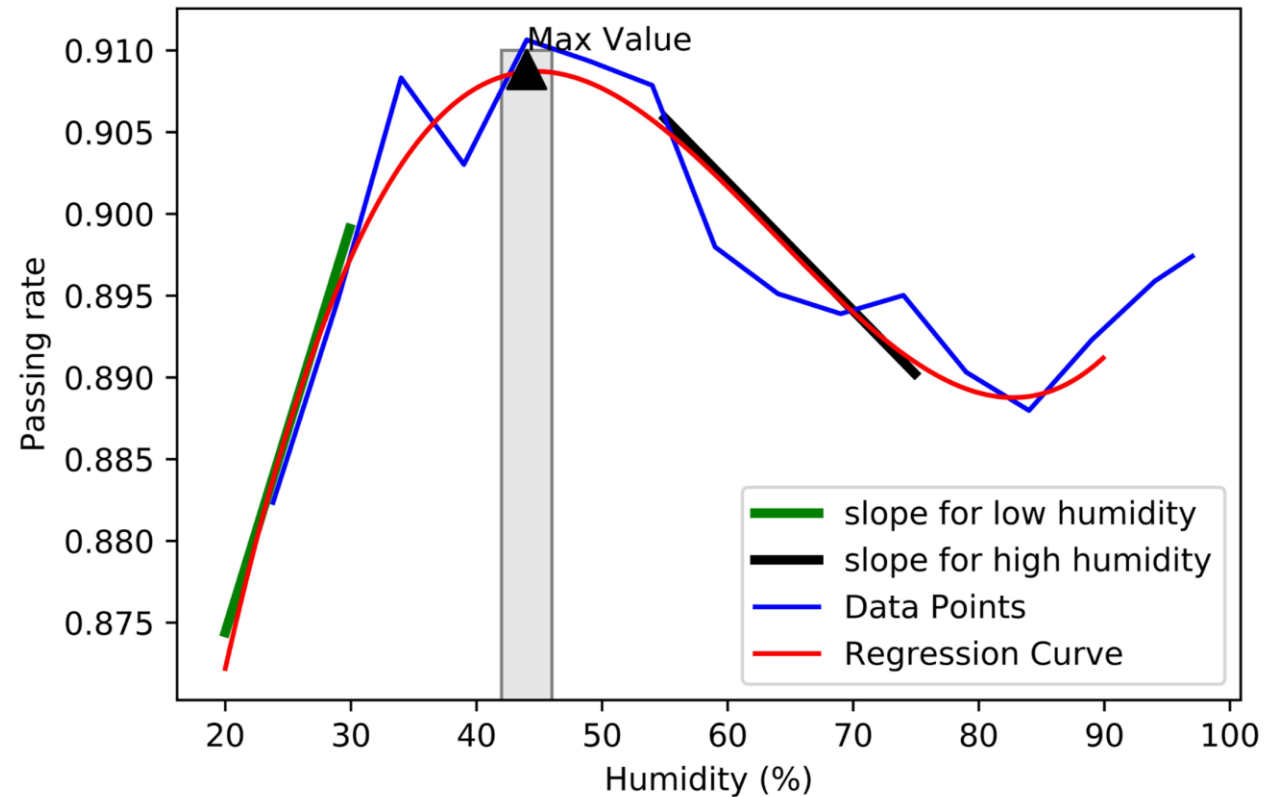
DATA SCIENCE CONSULTING FACULTY ASSOCIATES

- ▶ Welcome faculties with **data science & domain expertise** to serve as **faculty associates** in our data science consulting service.
- ▶ Integrate **data science** and **domain expertise** together.
- ▶ Provide **better & broader** consulting and training service to **corporate partners**.
- ▶ Team up for **data science related grant proposals**.
- ▶ Train students with both **data-science & domain expertise**.

DATA SCIENCE CONSULTING TRAINING & EDUCATION

- ▶ Training students with **real-world data** through consulting service
- ▶ **Vertically integrating** senior graduate students and undergraduate students for data science training
- ▶ Training students to use **large-scale GPU cluster** for **big data analysis** and **high-performance data processing**
- ▶ Training students with **data-science consulting, analytics, visualization, presentation & writing skills**
- ▶ Building **data-science learning communities**.

CASE STUDIES

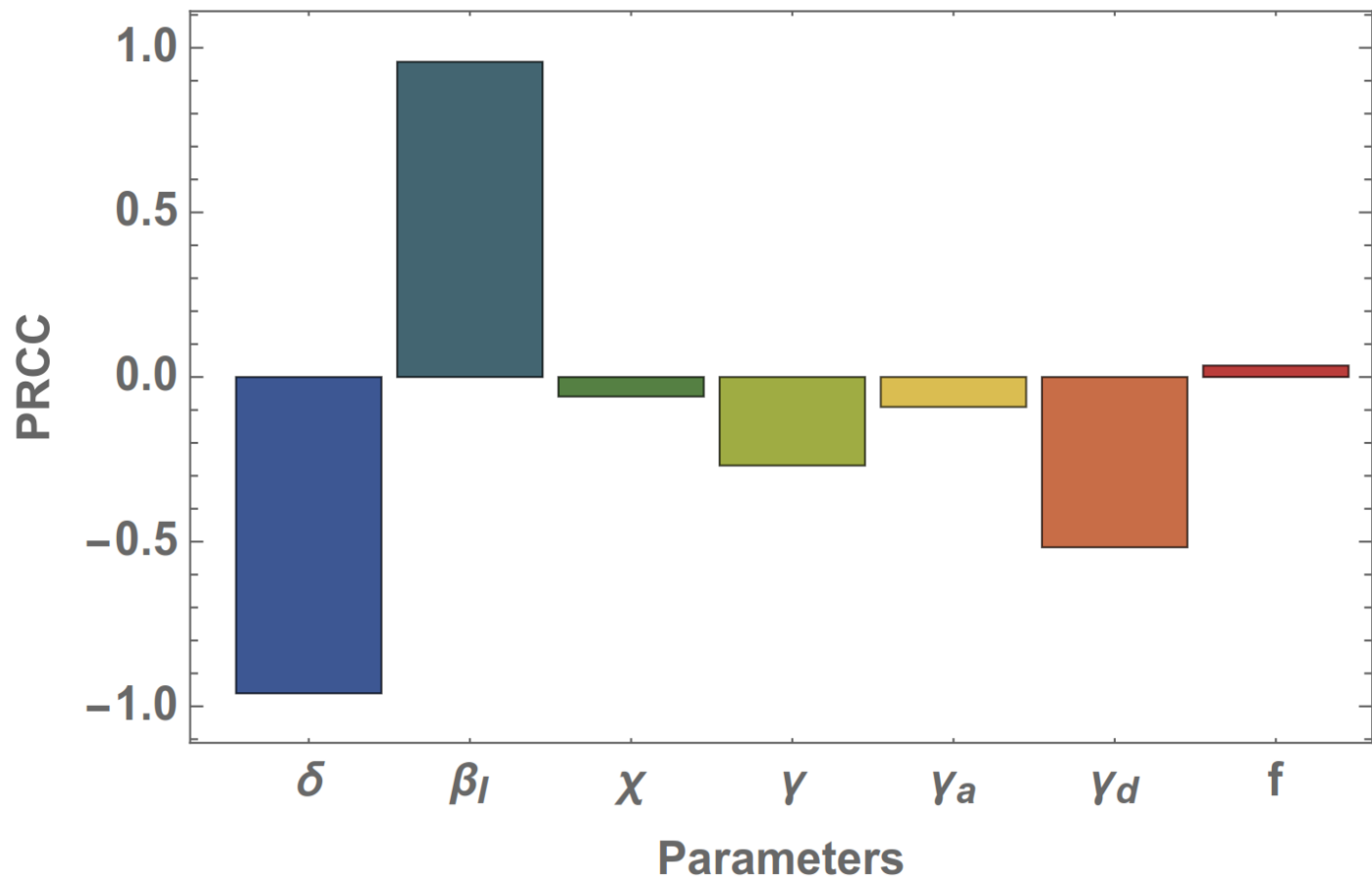


IMPROVING THE QUALITY OF CHRYSLER CROSSMEMBER CASTINGS

Summary:

A **crossmember** is a structural component that undergoes strict X-ray inspection to ensure its quality. It is critical to improving the **passing rate** in crossmember casting by **adjusting environmental and operational parameter settings**. By collaboration with Chrysler engineer, the optimal environmental and operational parameter settings are identified for making quality CHRYSLER crossmember castings through a novel machine learning algorithm. By employing the optimal setting, we can greatly increase the passing rate.

CASE STUDIES



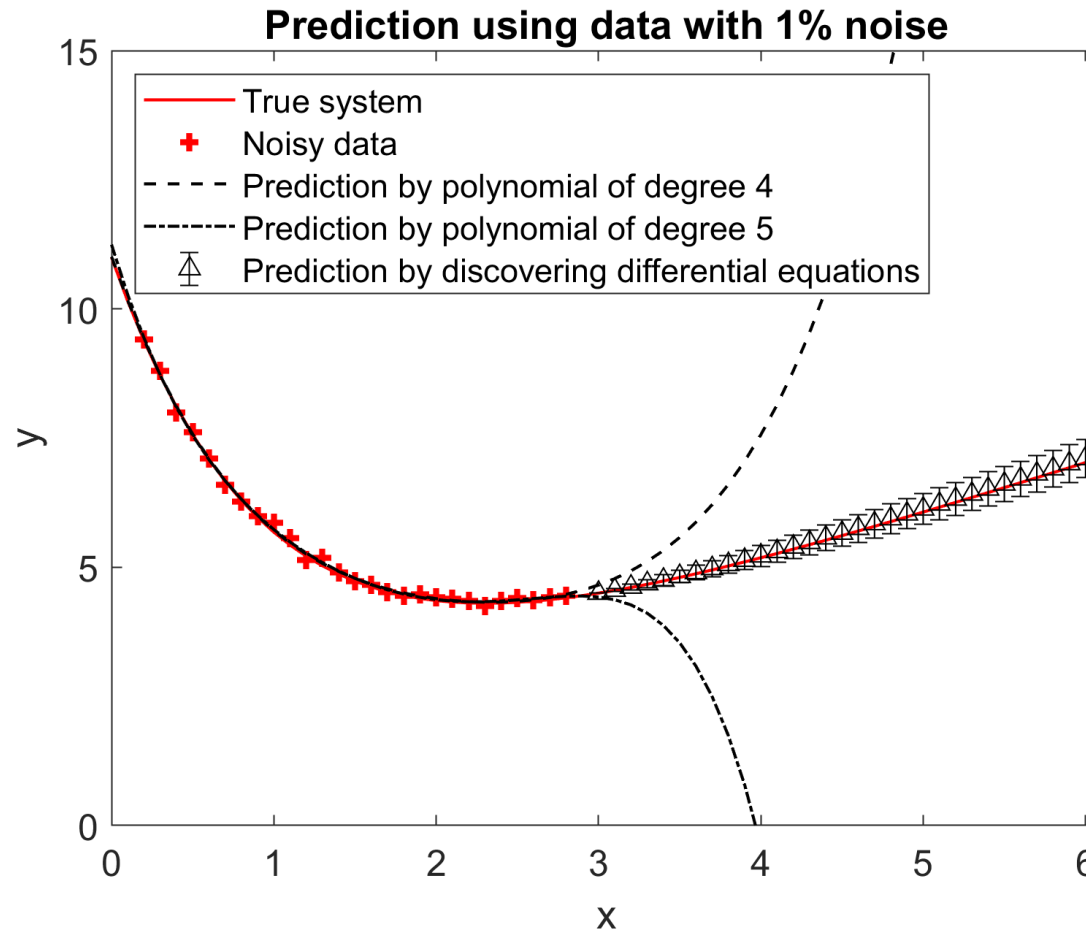
DESIGN OPTIMAL CONTROL STRATEGY FOR EBOLA OUTBREAK

Summary:

The 2014-15 **Ebola outbreak** in West Africa is a serious threat to global public health. To design and evaluate different control strategies for Ebola outbreak, we employ **machine learning, sensitivity analysis, & parameter estimation** to analyze the observation dataset. The results indicate that simultaneously **strengthening contact-tracing** and **effectiveness of isolation in the hospital** would be the most effective control strategies.

J. Ponce, Y. Zheng, **G. Lin**, Z. Feng, Assessing the effects of modeling the spectrum of clinical symptoms on the dynamics and control of Ebola, Journal theoretical biology, 467: 111-122, 2019.

CASE STUDIES



ROBUST DATA-DRIVEN DISCOVERY

Summary:

Can we automatically discover the hidden physical laws from noisy data? To solve this problem, we have developed a new machine learning approach on the **data-driven discovery of physical laws in implicit form from noisy datasets**. This approach is effective, robust and able to **quantify the uncertainties** by providing an **error bar** for each discovered candidate equations. This new method provides an effective way for the **robust data-driven discovery of physical laws**.

S. Zhang, **G. Lin**, Robust data-driven discovery of governing physical laws with error bars, Proceedings of the Royal Society of London. Series A, mathematical, physical and engineering sciences, in press, 2018. DOI: 10.1098/rspa.2018.0305

DATA SCIENCE RELATED FUNDING OPPORTUNITIES

[NSF Harnessing the Data Revolution \(HDR\): Institutes for Data-Intensive Research in Science and Engineering - Ideas Labs \(I-DIRSE-IL\)](#)

Deadline: Monday, March 4, 2019

[DOE Data Science for Discovery in Chemical and Materials Sciences](#)

Deadline: Friday, March 8, 2019

[NSF Harnessing the Data Revolution \(HDR\): Transdisciplinary Research in Principles of Data Science Phase I](#)

Deadline: Monday, March 25, 2019

[NSF Harnessing the Data Revolution \(HDR\): Institutes for Data-Intensive Research in Science and Engineering - Frameworks \(I-DIRSE-FW\)](#)

Deadline: Tuesday, May 7, 2019

For more funding opportunities, Go to IDSI website: Research -> [Funding opportunities](#)

Any Questions?

[HTTPS://WWW.PURDUE.EDU/DATA-SCIENCE/DSCS/](https://www.purdue.edu/data-science/dscs/)

Contact: Guang Lin, guanglin@purdue.edu

Special thanks go to: Patrick Wolfe, Sunil K. Prabhakar, Pankaj Sharma, Daniel E. Hirleman, Bruce Craig, George McCabe, Greg Buzzard, Hao Zhang, Larry S. Sommers, Preston M. Smith, Adrienne N. Thompson, Holly G. Graef, and Pamela G. Burroff-Murr.