# APPLYING SOPHISTICATED ANALYTICS TO ACCELERATOR DATA AT BNL'S COLLIDER-ACCELERATOR COMPLEX:
## BRIDGING TO REPOSITORIES, TOOLS OF CHOICE, AND APPLICATIONS*

K.A. Brown, P. Chitnis, T.D'Ottavio, J.Morris, S.Nemesure, D.Thomas

C-AD Dept., BNL, Upton, NY

BROOKHAVEN NATIONAL LABORATORY

CONTROLS
Collider - Accelerator Department

---

## What is RHIC?

- RHIC = Relativistic Heavy Ion Collider: Discovered the Quark-Gluon Plasma (QGP) & its Perfect Liquid property
- Only Polarized Proton Collider in the world. Solving the "spin puzzle" or how the quarks and gluons contribute to the protons spin
- In the future will host eRHIC, an electron ion collider.

## RHIC Controls: Data Collection and Storage
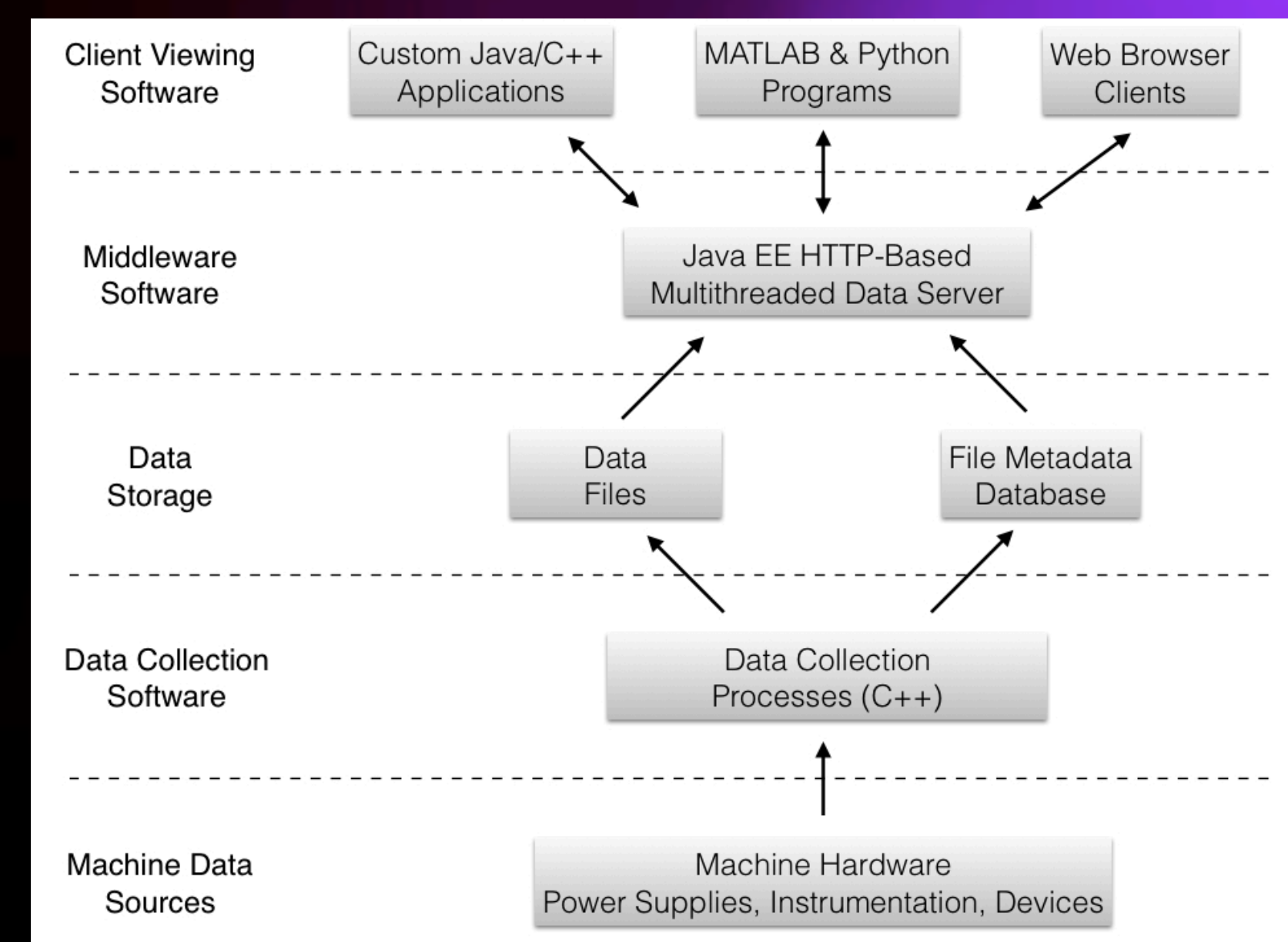


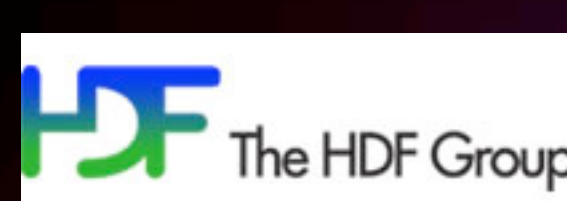Above: Three tier data layers for Accelerator Data.
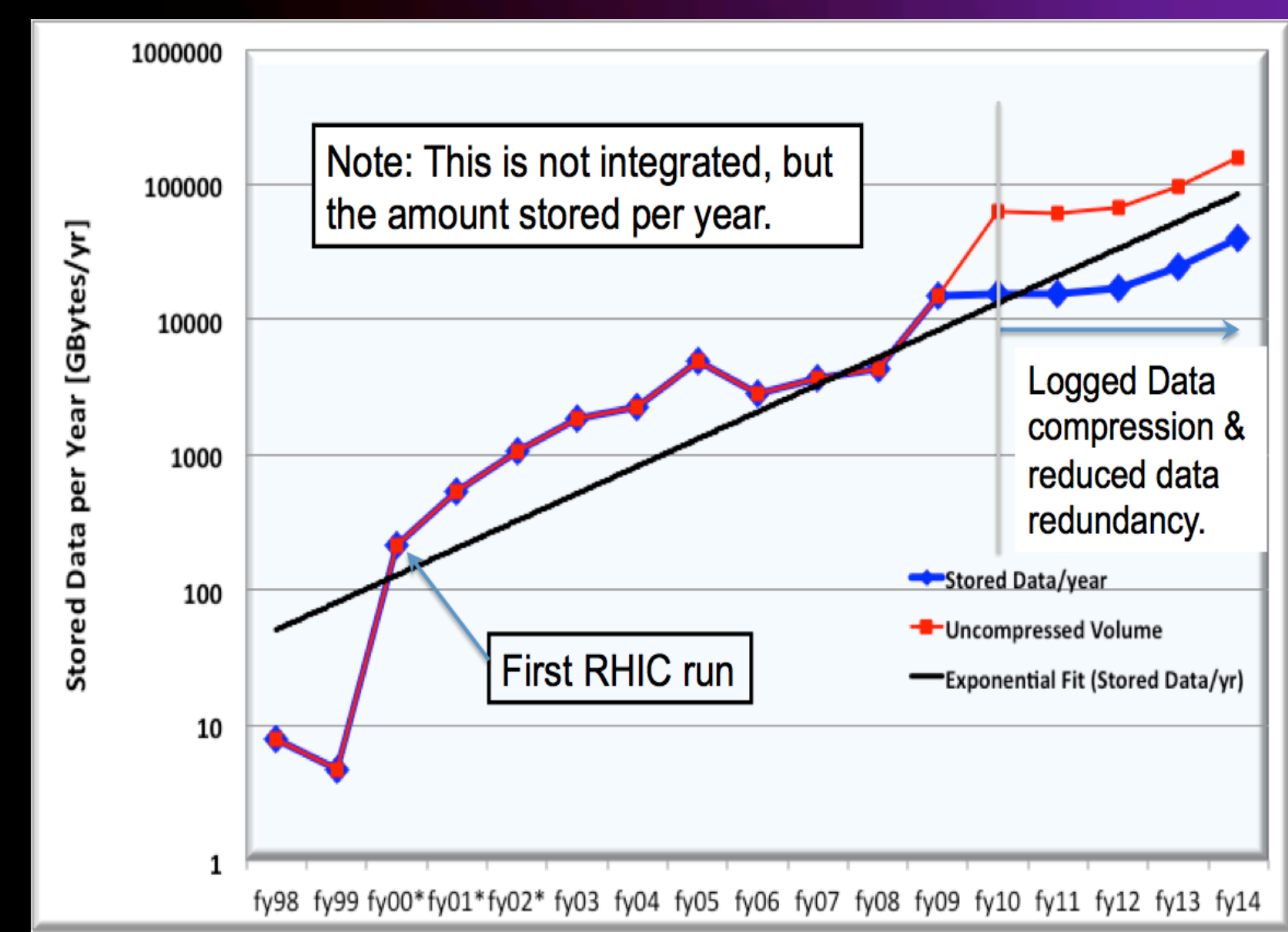


Above: Amount of data stored per year over the past 15 years has increased exponentially. The data stored/year in 2002 was a little over 1 TB, while in 2013 it was over 100 TB. Logged Data compression began in 2010, reducing the data written to disk by a factor of 4. Come 2024 we may be storing 2 PT/year!

### Background Image:
*Elliptic and triangular flow in event-by-event (3+1)D viscous hydrodynamics*
Bjoern Schenke, Sangyong Jeon, Charles Gale
Phys.Rev.Lett.106:042301.2011
http://arxiv.org/abs/1009.3244

---

## MODERN TOOLS

HDF The HDF Group

The Apache Software Foundation
Community-led development since 1999.

"HDF5 is a unique technology suite that makes possible the management of extremely large and complex data collections."
https://www.hdfgroup.org/HDF5/whatishdf5.html

hadoop

Spark

cassandra

mongoDB

Distributed Database systems:
- Apache Hadoop is a "framework" for distributed processing of large data sets across multiple computer systems.
- Apache Cassandra is a scalable distributed NoSQL database, designed for managing large amounts of data across multiple systems.
- Apache Spark is a "processing engine" designed to be used with Hadoop, Cassandra, and other systems for processing streaming data, interactive queries, and machine learning.
- MongoDB is another scalable database designed for large data sets and ease of development.

MathWorks    SciPy.org    ENTHOUGHT

R

MATLAB is a powerful system for applying many mathematical algorithms for the analysis of accelerator data and is used by many facilities around the world.
http://www.mathworks.com/products/matlab/

R is a free software project aimed at statistical computing and visualization.
https://www.r-project.org/

SciPy is an open source collection of Python based tools for math, science, and engineering.
http://www.scipy.org/

rapidminer    Weka    orange

### Other useful tools:
Rapidminer : predictive analytics tools
WEKA: machine learning
Orange: data visualization and analysis
KNIME: open analytics platform & informatics tools
NLTK & Apache OpenNLP: natural language toolkits

Plus many more
https://github.com/onurakpolat/awesome-bigdata

---

## EXAMPLES

### MATLAB Controls Data Viewer
- Data Server is http protocol server
- Data in XML format
- MATLAB urlread/urlwrite can be used to access
- Xmlreadstring() returns DOM node
- Use standard DOM package to collect data
- Data sets packaged into MATLAB structure cells, to retain controls names space on data imported into MATLAB environment.
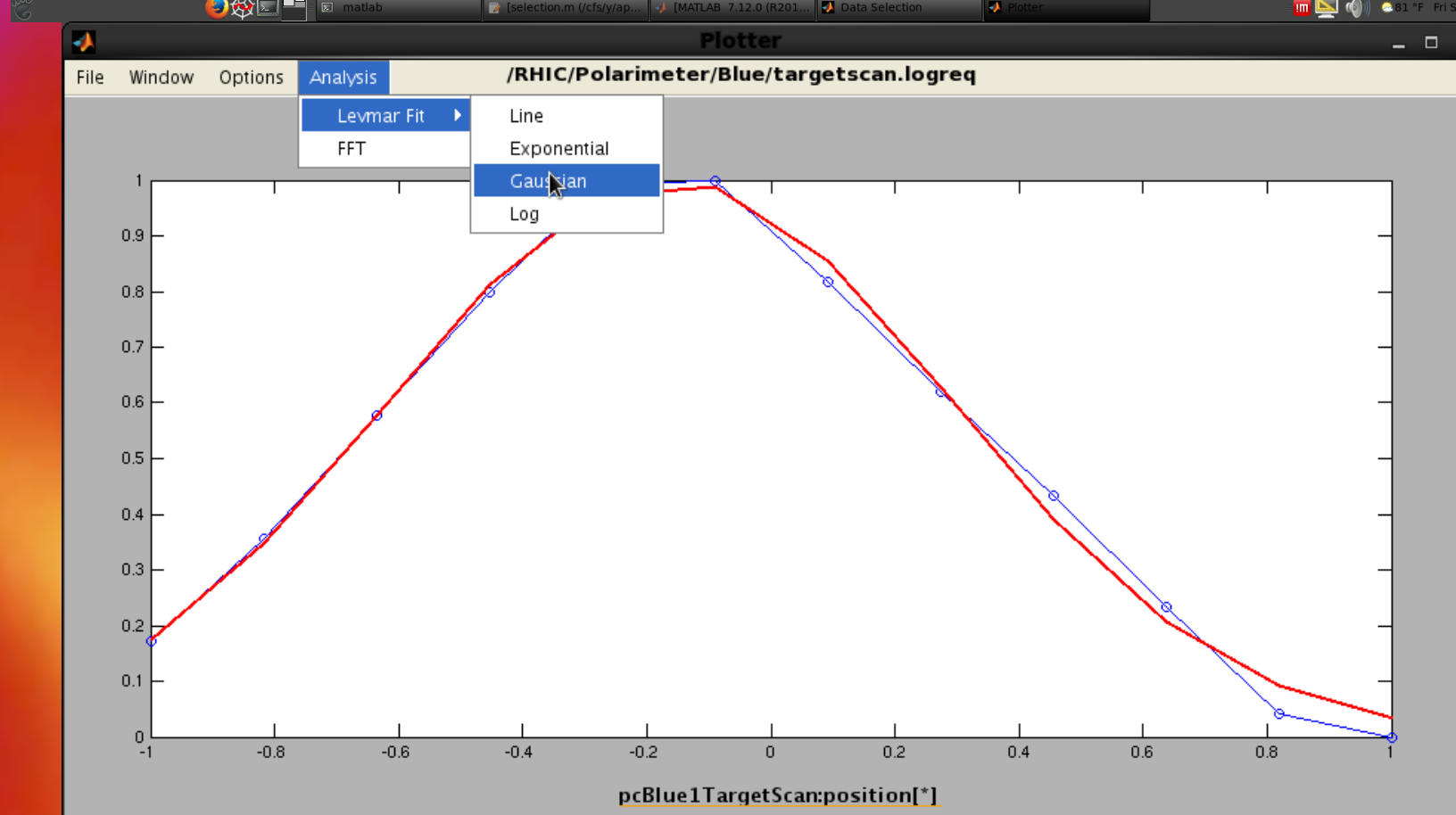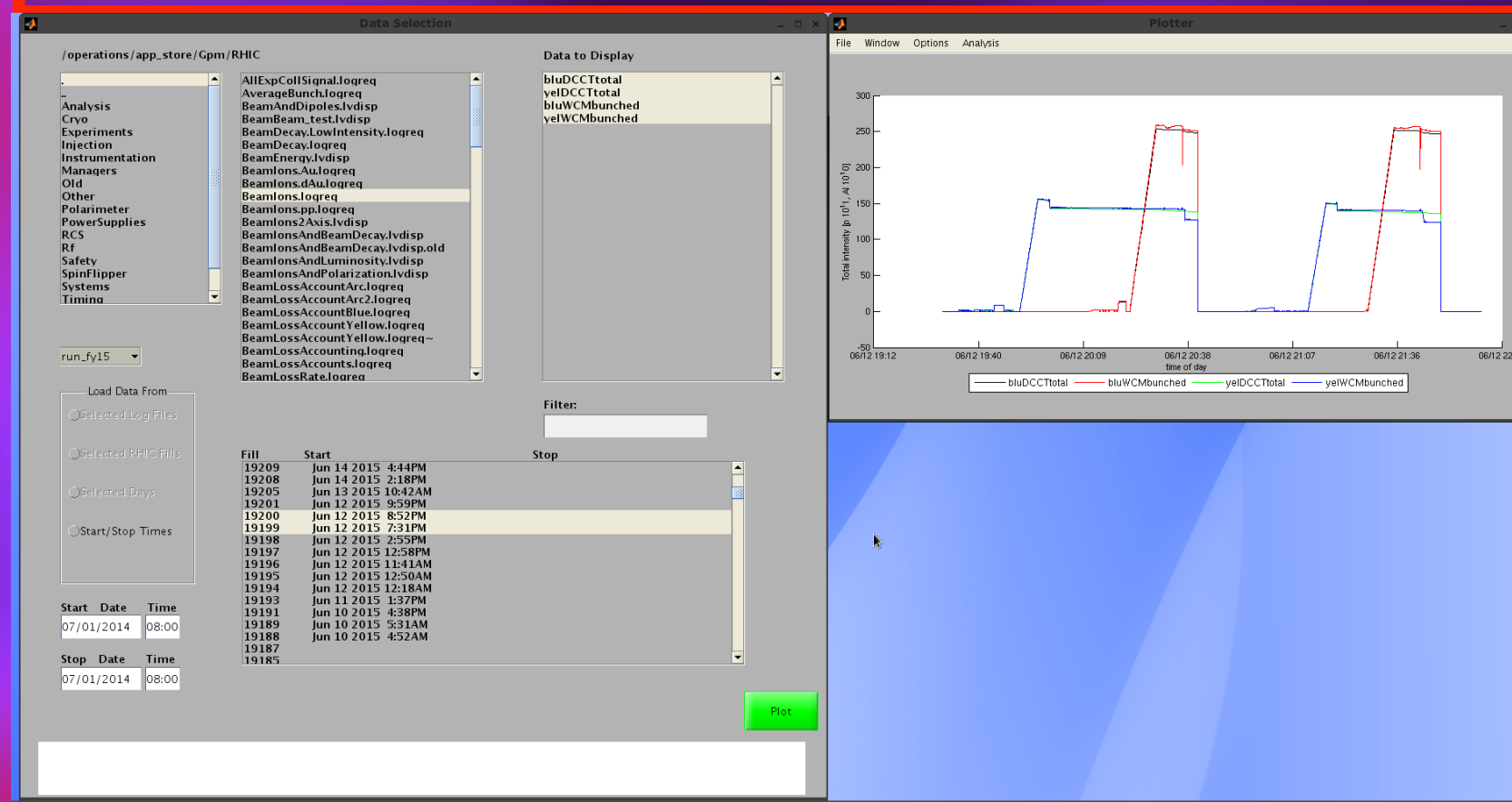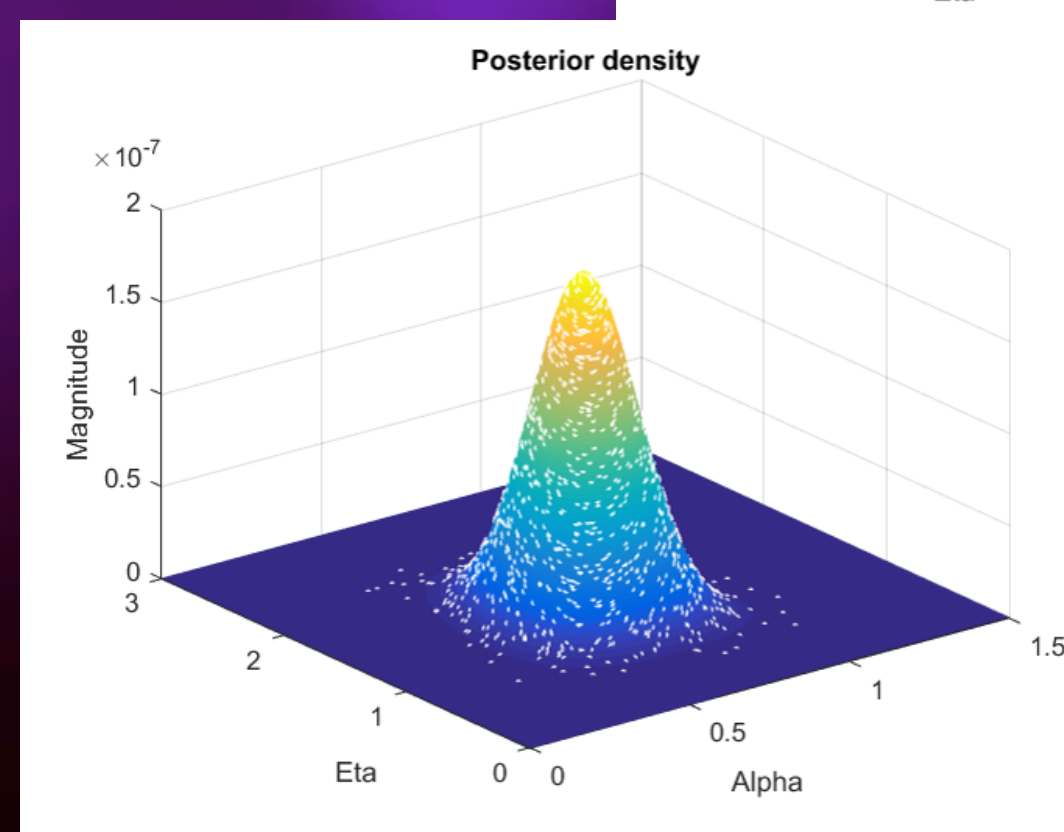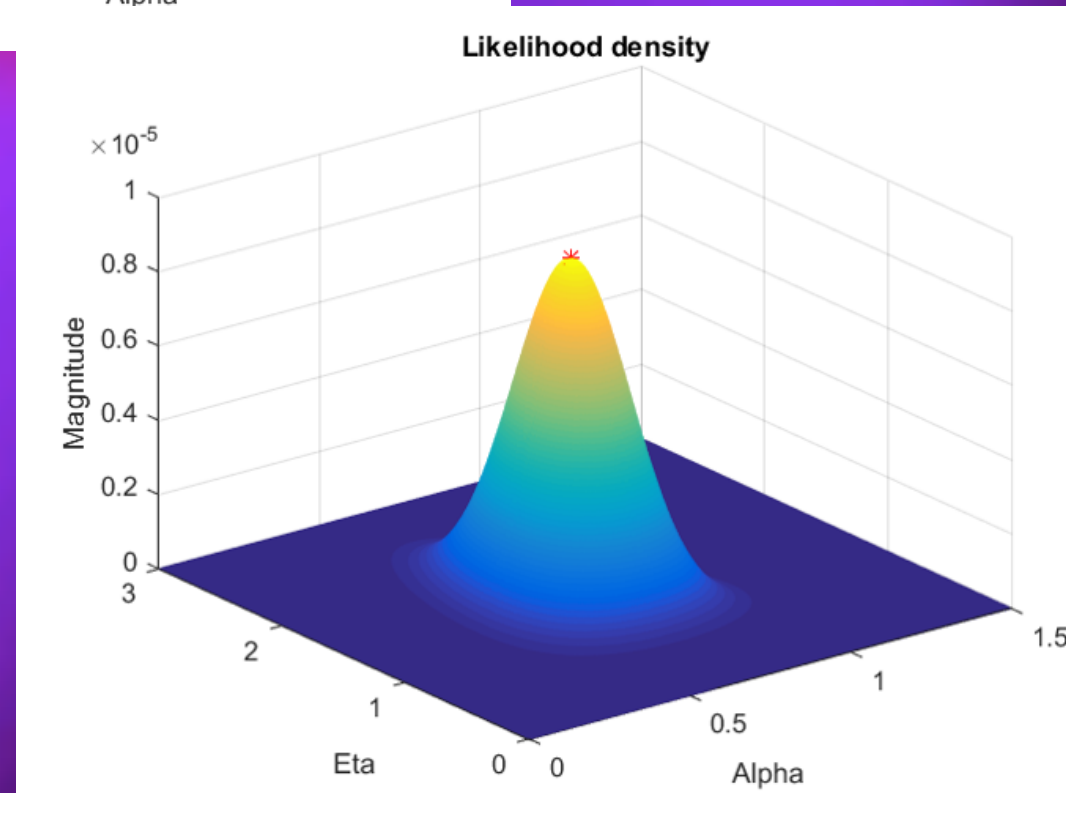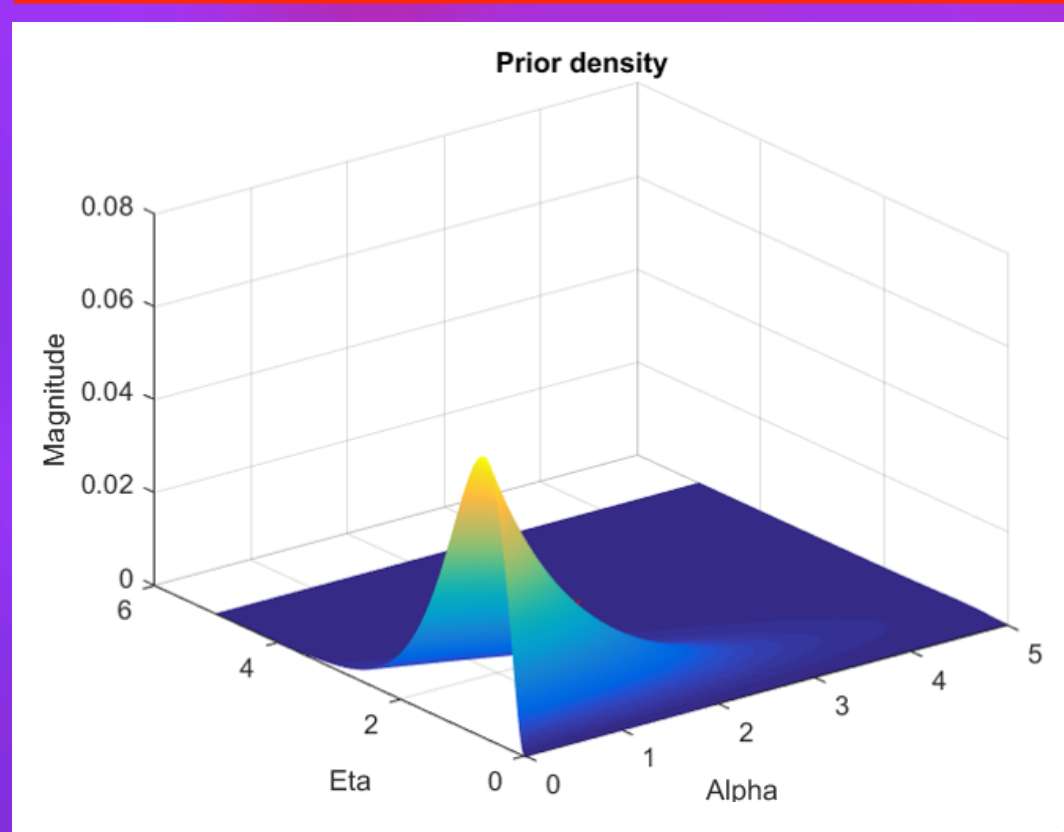


### Analyzing Machine Protection Data

In R & MATLAB can build models and analyze data
- Bayesian analysis
- 2-parameter Weibull survival model
- Markov Chain Monte Carlo methodology
- Metropolis-Hastings algorithm

* See " MOD3I01 Bayesian Reliability Model for Beam Permit System of RHIC at BNL"



### Better Quench Detector Tables

RHIC Quench detectors rely on a simple electrical model. That model has limitations that can lead to false quench events and possible (momentary) blind conditions.
- Non-linear electrical behavior
- Include eddy current components
- Include parasitic capacitances
- These are on top of the normal inductive and lead resistance model traditionally used
- Use piecewise regression to examine saturation effects
- comprehensive residual analysis

* See MOM310 "Nonlinear System Identification of Superconducting Magnets of RHIC at BNL"
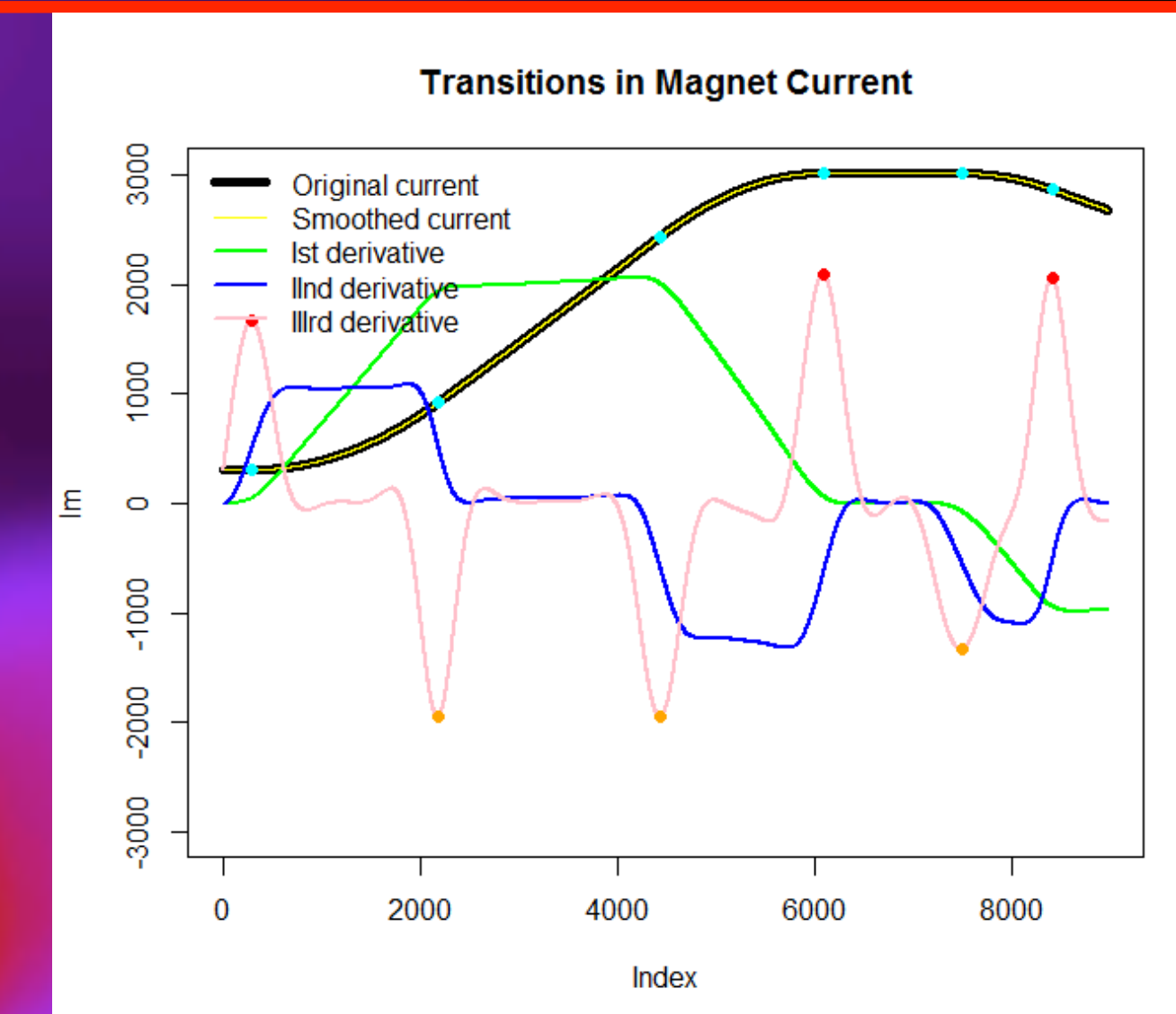


Above: Example of using R to develop an improved model to fit inductance data used in quench detector systems in RHIC.
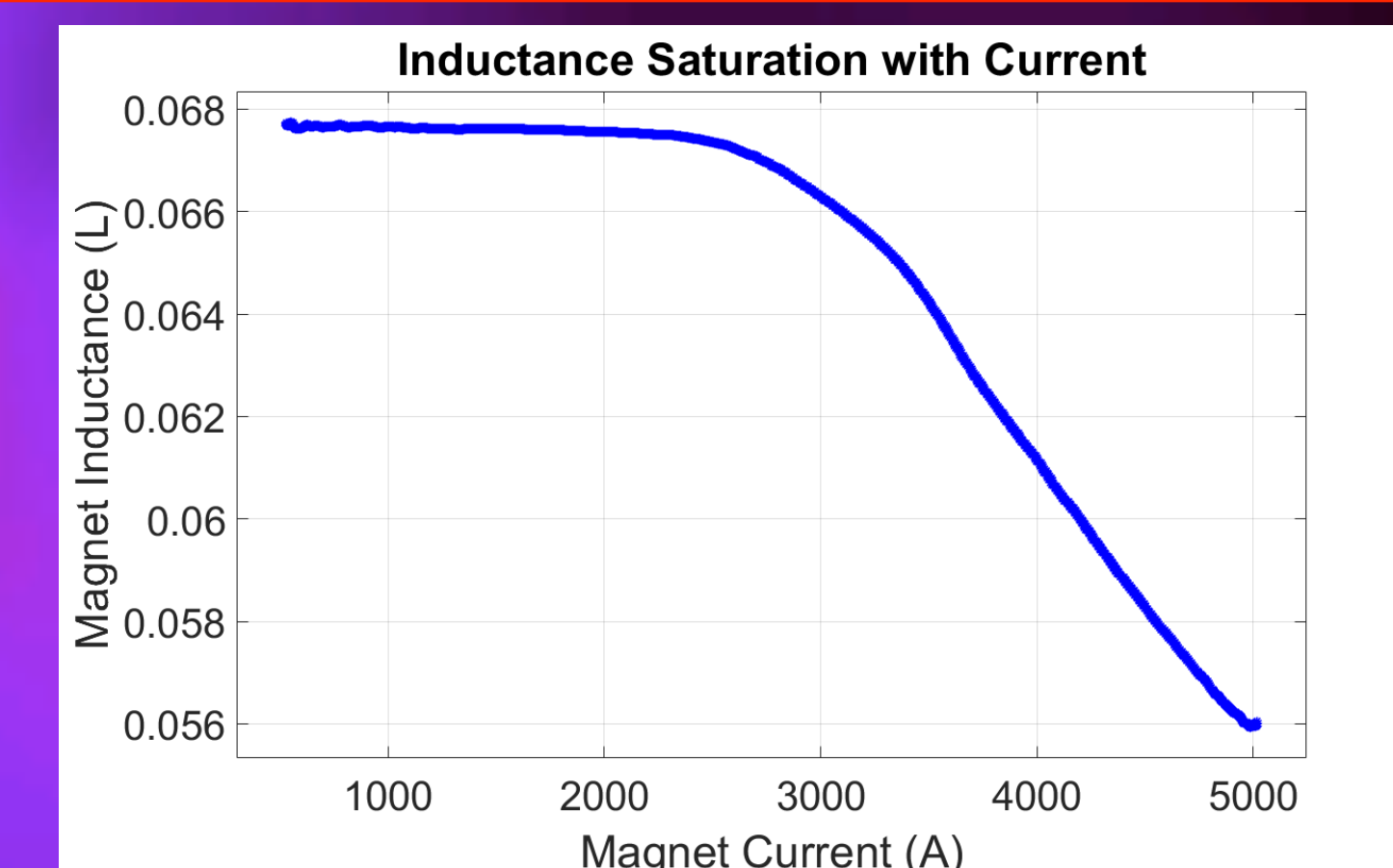


### Neural Network & Markov Models

Use of NumPy and SciPy to do predictive analytics
- neural network analysis = reinforcement learning in the form of a Markov decision process
- optimal minimization cost using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm
- Use Levenberg-Marquardt for back-propagation fitting
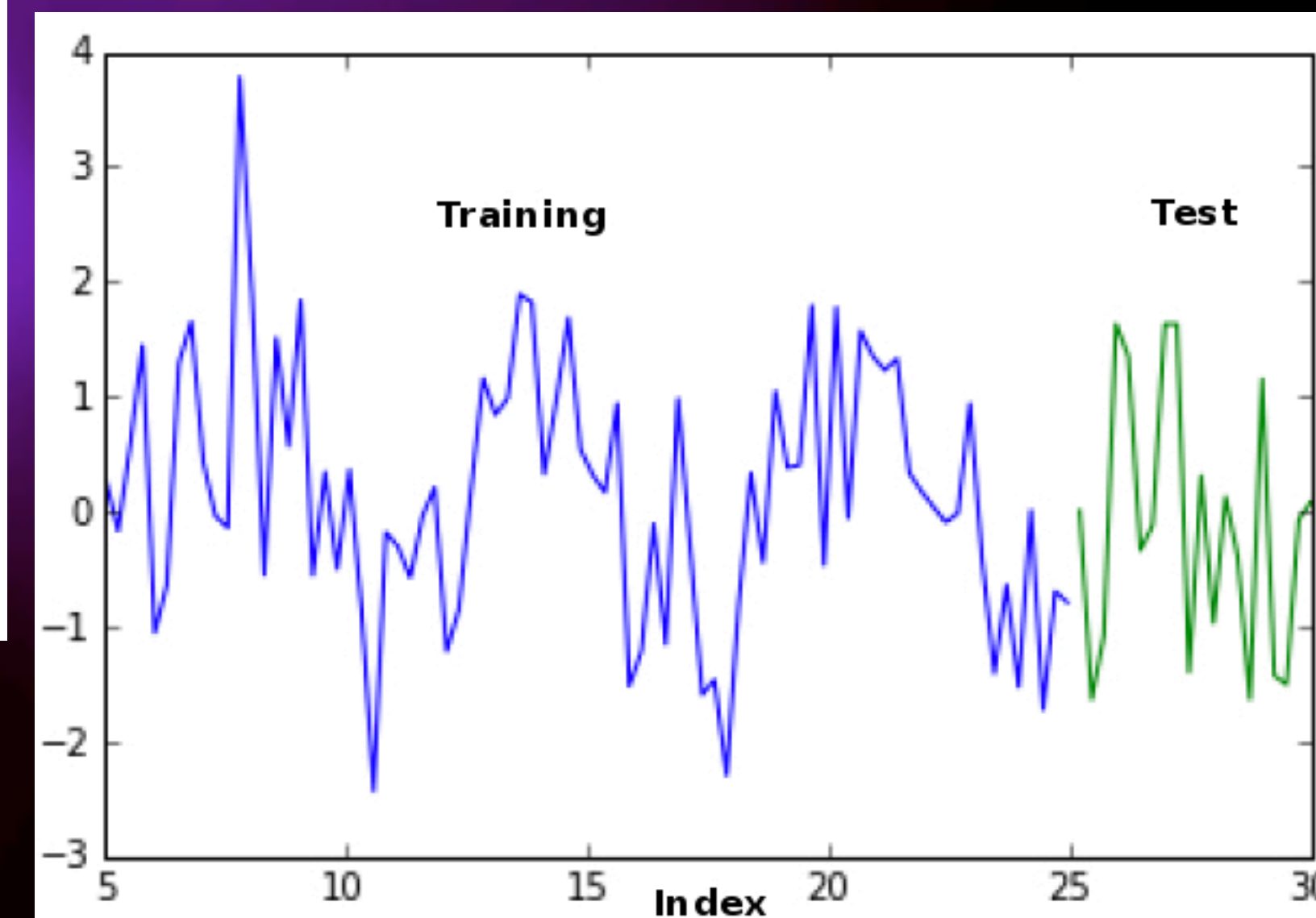


Above: An example of predictive analysis using a neural network derived model of past data.