# Capturing Operational Value Through Data Analytics

**Doug White**
**Principal Consultant**
**Emerson Process Management**
**Houston, Texas**

# Speaker

**Doug White (doug.white@emerson.com)**

**Principal Consultant**

**PlantWeb Solutions Group**

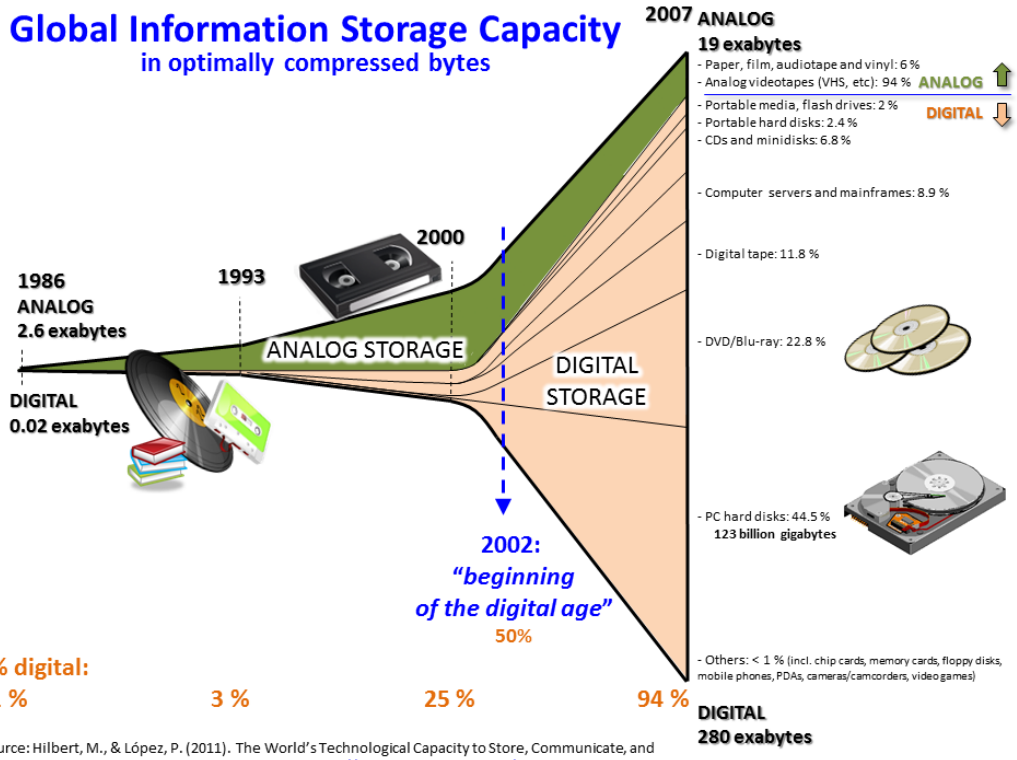
**Emerson Process Management**

**Background: Many years experience designing, justifying, installing and commissioning advanced real time computer/ automation applications in the process industries.**

EMERSON
Process Management

AIChE
Southwest Process Technology Conference

# Data is Increasing




Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

**Exponential growth of global storage capacity and processing power**

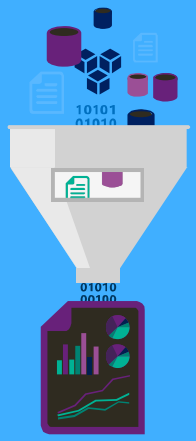
# Data Is Increasing

- Walmart's Online Database
  - More than 250 billion rows
  - More than 20 million updates per hour
  - 3 years history available online
  - 24,000 queries per hour from 1500 users
- Internet Devices
  - Gardner group estimates 5 billion connected devices in 2015 – 25 billion by 2020
  - Plant device connectivity is increasing similarly
- Process Company
  - 10 million tags across 15 sites
  - Implementing system to provide real time access to 3 years of 1 minute sampled historical data for tags
- Refining Company
  - Reported total of 80 billion data items from four sites in one year

# Types of Data Analytics

**Retrospective analytics**


**Real-time analytics**




**Predictive analytics**

**Intelligent SaaS apps**

Source: EdX Course: DAT203x: Data Science and Machine Learning Essentials, Ch.2; Microsoft

AIChE Southwest Process Technology Conference

# Creating Value From Data - Predictive Analytics – Looking Forward

## Looking Back

What happened and how did it compare to plans?

Gaps? Why?

What is likely to happen in the future?

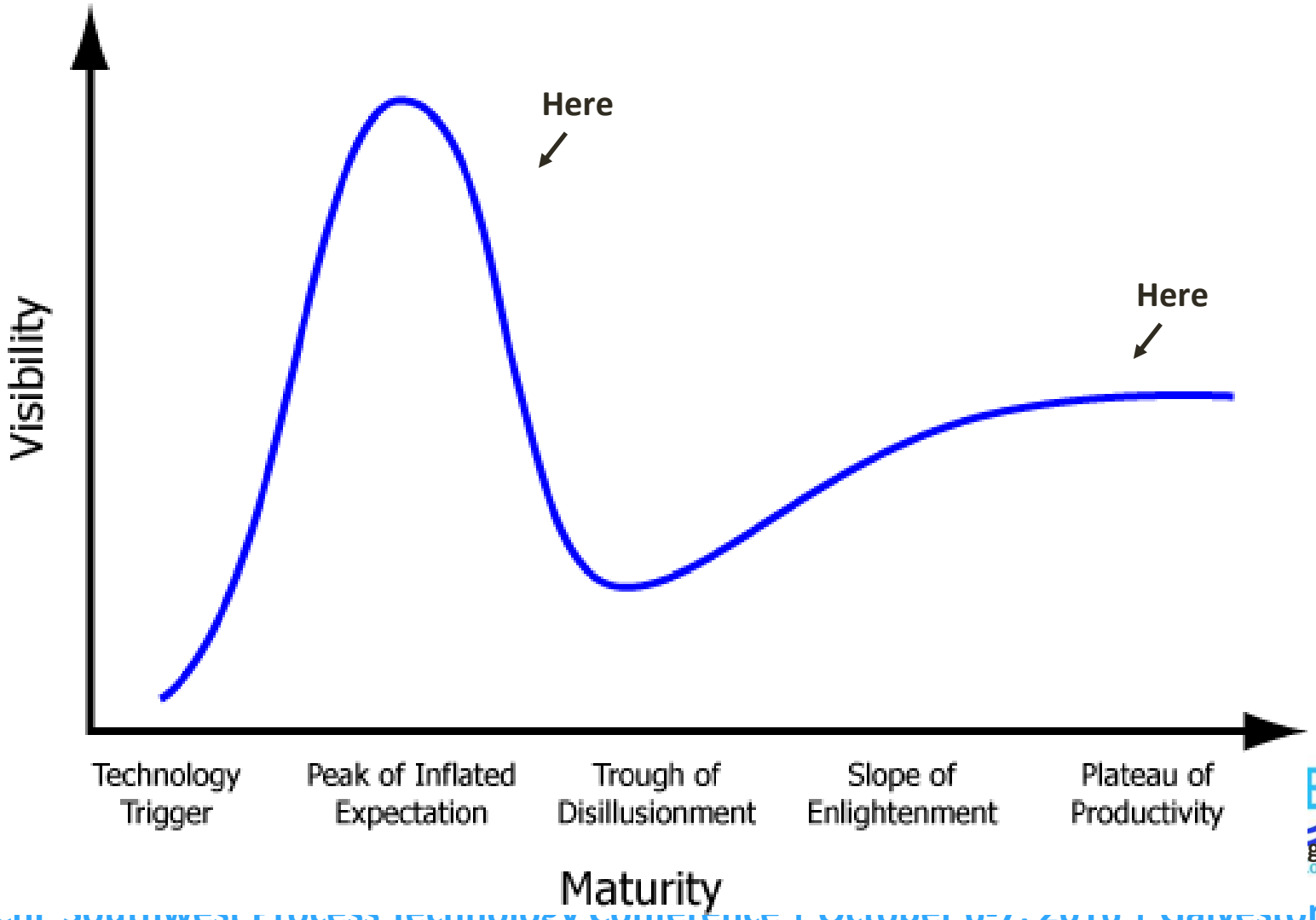
Fix or Mitigate Problems

## Looking Forward

**Predictive analytics, in general, is the use of current and historical data in conjunction with system models to predict future trends in real-time (near).**

AIChE
Southwest Process Technology Conference

# Where is Predictive Analytics in the "Hype" Evolution?



# Predictive Analytics - Moneyball

- In 2001, the Oakland A's (salaries $44m) lost to the New York Yankees (salaries $125m) in the divisional playoffs and then lost 3 top players to free agency.
- Billy Beane, the manager, hires Peter Brand, a new graduate, to implement statistical analysis of previous seasons to determine what it takes to win. He develops correlations that go against some of the traditional baseball ideas. Beane supports using the approach to pick new players.
- In 2002, The A's set a league record by winning 20 games in a row on the way to winning 95 games overall and making the league playoffs – and doing this again in 2003.

# Predictive Analytics - Predicting Supreme Court Decisions

- **Andrew Martin, a professor of political science at Washington University, developed a statistical model of supreme court decisions based on actual decisions from 1994 to 2001 (no change in justices during that period.**
- **He used the model to predict decisions for the 2002 term. Results were compared against a panel of distinguished legal individuals experienced with the supreme court.**
- **For the 68 cases in the 2002 term:**
  - **Model accuracy – 75%**
  - **Experts accuracy – 59%**

AFPM
American
Fuel & Petrochemical
Manufacturers

# How Does This Apply (And Not Apply) To The Process Industries?

# Plant ("Near") Real Time Data Includes Many Types

| Numeric Data | Spectral Data | Video/ Images | Text |
|---|---|---|---|
| Process/ Offsites/ Utilities/ Terminal Data | Analyzers – NIR, etc | Cameras – Flares, Process, Security, Etc | Daily Plans, Notes, etc |
| Lab Data | Rotating Machinery – Vibration, etc. | Thermography | Maintenance - Inspection, Work Orders, etc |
| Accounting, KPI's, Planning, Scheduling, etc | Infrared/ Acoustic/ Ultrasonic – Event Detection | Misc | E-Mail |

# Typical Refinery DCS IO (Input / Output) Numbers – Continued Increase

1990 – "Hard" IO
Count: 20,000

2015 – "Hard" IO
Count: 50,000

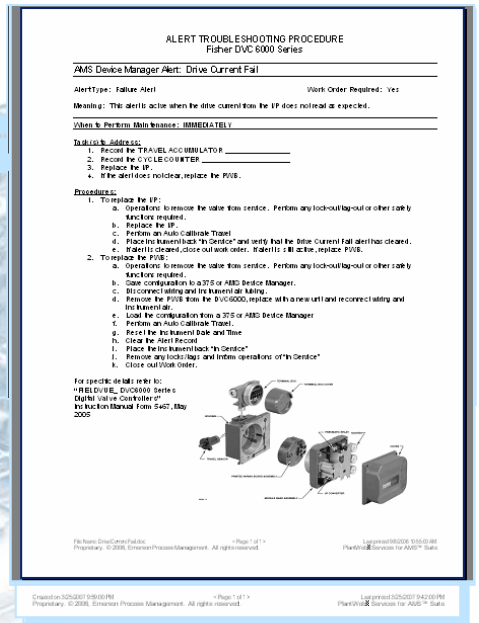
2025 – "Hard" IO
Count: 100,000+

Timeline

Photo Source: US DOE

# Remote Access - “Internet of Things” in Plants

SERIAL CARD
PRODUCT DATA SHEET
MANUAL
INSTALLATION
CALIBRATION
QUICK START GUIDE
ALERT REFERENCE
DATA CAPTURE
VISUAL INSPECTION
USER CONFIGURATION
AND MORE…



ALERT TROUBLESHOOTING PROCEDURE
Fisher DVC 6000 Series

AMS Device Manager Alert: Drive Current Fail

AlertType: Failure Alert — Work Order Required: Yes

Meaning: This alert is active when the drive current from the I/P does not read as expected.

When to Perform Maintenance: IMMEDIATELY

Tasks to Address:
1. Record the TRAVEL ACCUMULATOR _____
2. Record the CYCLE COUNTER _____
3. Replace the I/P.
4. If the alert does not clear, replace the PWB.

Procedure:
1. To replace the I/P:
   a. Operations: to remove the valve from service. Perform any lock-out/tag-out or other safety functions required.
   b. Replace the I/P.
   c. Perform an Auto Calibrate Travel
   d. Place the Instrument back "In Service" and verify that the Drive Current Fail alert has cleared.
   e. If alert is cleared, close out work order. If alert is still active, replace PWB.
2. To replace the PWB:
   a. Operations: to remove the valve from service. Perform any lock-out/tag-out or other safety functions required.
   b. Save configuration to a 375 or AMS Device Manager.
   c. Disconnect tubing and the Instrument air tubing.
   d. Remove the PWB from the DVC6000, replace with a new unit and reconnect tubing and the Instrument air.
   e. Load the configuration from a 375 or AMS Device Manager
   f. Perform an Auto Calibrate Travel.
   g. Reset the Instrument Date and Time
   h. Clear the Alert Record
   i. Place the Instrument back "In Service"
   j. Remove any locks/tags and inform operations of "In Service"
   k. Close out Work Order.

For specific detail refer to:
"FISHER, DVC6000 Series Digital Valve Controller" Instruction Manual Form 5467, May 2005

**Available Anywhere**

**Photo Ref: Emerson Process Management**

***Move the data to the expert, not the expert to the data!***

AFPM
American
Fuel & Petrochemical
Manufacturers

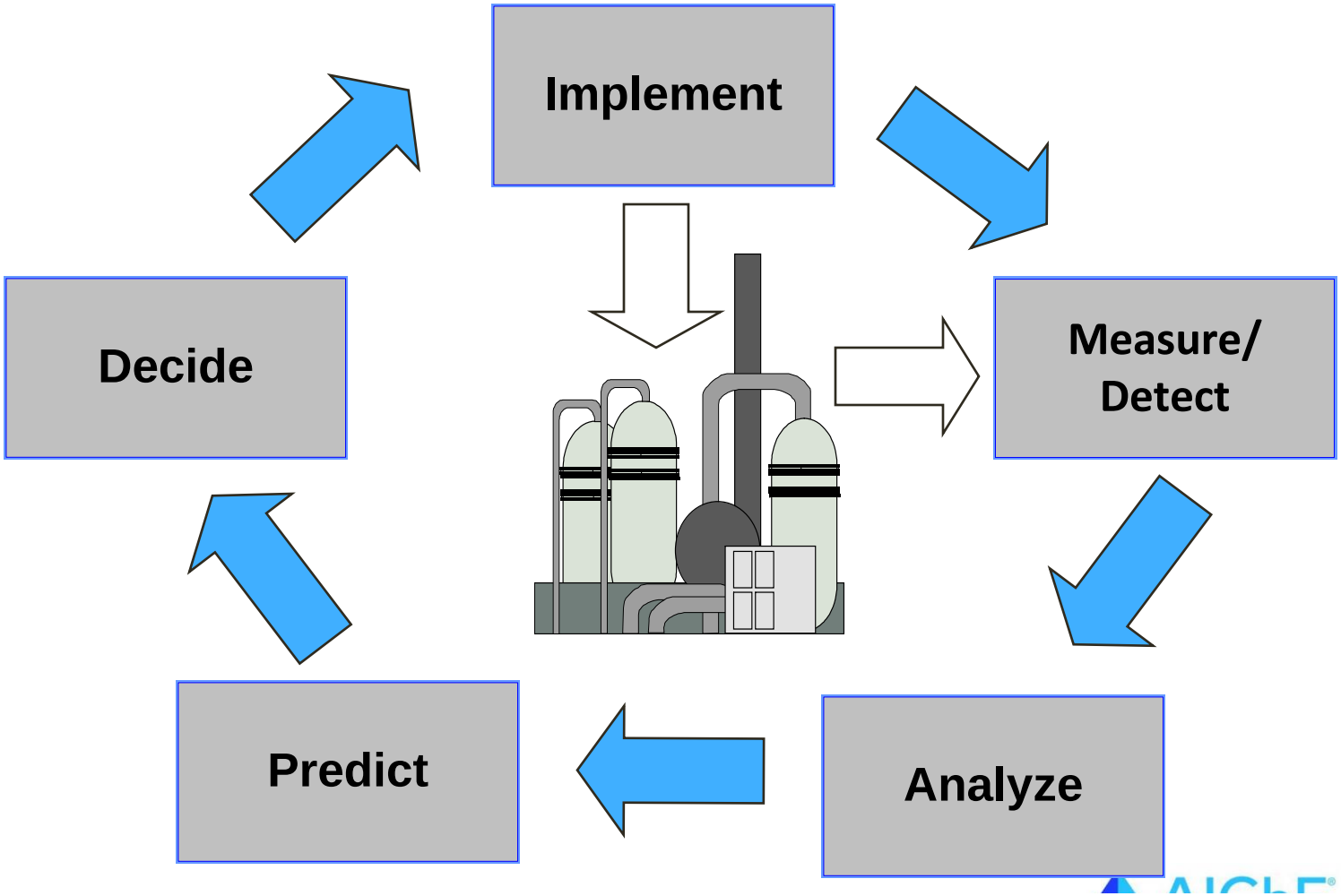
# Making Sense And Cents From Data - Where Is The Payback?

# Plant Operating Objectives - The Four Zero's

- ***Safety*** **– the goal is zero serious safety incidents**
- ***Sustainability*** **– the goal is zero significant environmental incidents, excess energy use and excess waste**
- ***Reliability*** **– the goal is zero unscheduled downtime**
- ***Financial*** **- the goal is zero lost profit opportunities**

***How can Predictive Analytics support these objectives?***

# Plant Decision Cycle

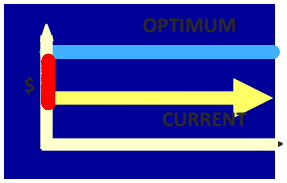



***To have a financial impact predictive analytics has to improve this cycle – reduce delays, reduce uncertainty, etc.!***

# What Impact Can Predictive Analytics Have?

- **Safety**
  - **Avoiding incidents through early detection of potential hazardous situations**
- **Availability/ Reliability**
  - **Anomaly detection – identifying precursor events to unscheduled equipment outage or problems**
  - **Performance monitoring – detecting loss of process/ equipment performance before it impacts production capacity**
- **Sustainability**
  - **Comparing current usage of resources such as energy to its expected usage under current conditions and determining possible causes of variation**
- **Financial Optimization**
  - **Detecting and dissecting complex interacting constraints on production**
  - **Determining reasons for product quality/ yield issues**
  - **Understanding patterns and relationships – developing statistical models that explain them**

# What are the Components of a Predictive Analytics Implementation?



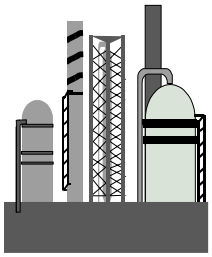


**Acquire:**
Dynamic Event Detection and Sensor Data Capture

**Analyze:**
Extract Information, Patterns

**Prediction:**
Models

**Present:**
HMI, Distribution



**Action:**
Algorithms

Photo Ref: Emerson Process Management

***Move from reaction to events to anticipation and optimization***

# Major Predictive Analytics Techniques

- **Numerical Data (with adjustment for time series dependence)**
  - **Develop Predictive Model**
    - **Linear Models**
      - **Continuous Variables – *Generalized Linear Regression***
      - **Binary Outcome (Yes/No) – *Logistics Regression***
      - **Discrete Outcome (1,2,3,4,..) *Classification and Reference Trees (CART + Random Forest)***
    - **Non-Linear Models**
      - **Continuous Variables - *Neural Nets***
      - **Discrete Outcome (1,2,3,4,..) - *(CART + Random Forest)***
  - **Detect Anomaly**
    - ***Principal Component Analysis - PCA***
- **Textual Data**
  - ***Textual Analytics***
- **Spectral Data**
  - ***Nearest Neighbor***
  - ***Regression***
- **Video/ Images**
  - ***Neural Nets***

AFPM
American
Fuel & Petrochemical
Manufacturers

# Typical Analytics Project

# Typical Analytics Project Steps

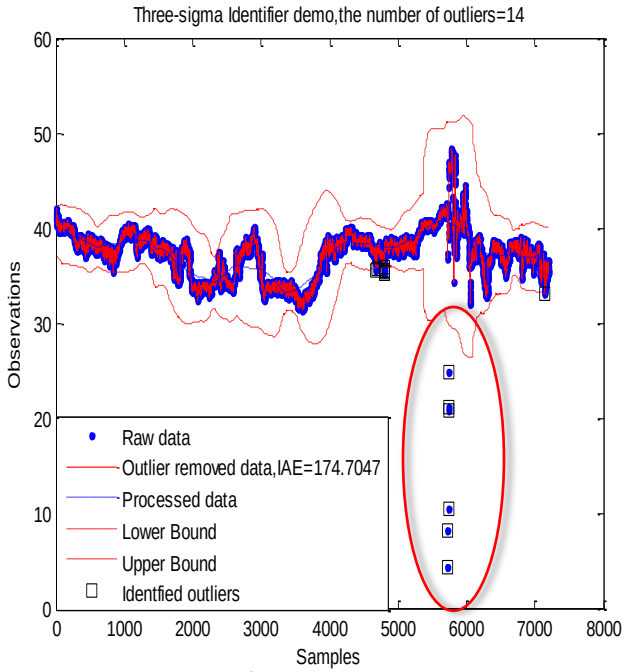
1. **Problem Definition – How will results be used to improve business performance**
2. **Data Location, Interpretation and Acquis**
3. **Data Cleaning, Consolidation and Transformation**
4. **Model Building and Evaluation**
5. **Model Deployment, Monitoring and Updates**

**Often more than 50% of project hours**

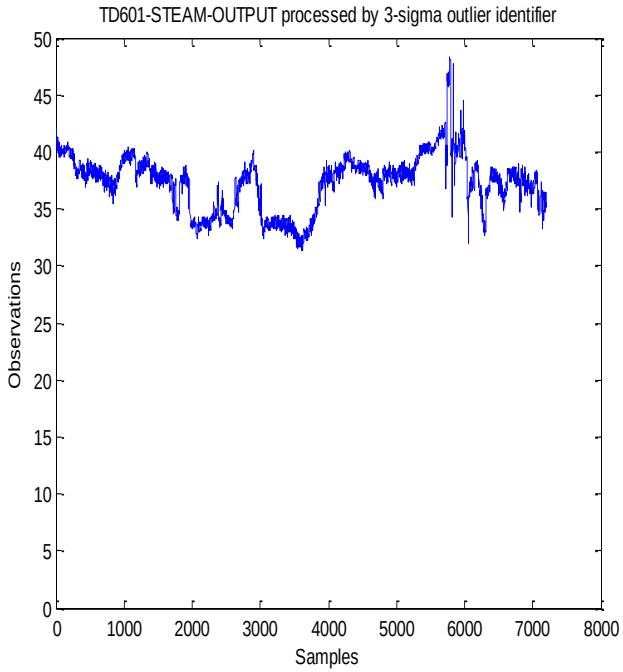
# Plant Data Typical Questions

- **Data**
  - What data should we collect?
  - How should we store it? Particularly non-numeric data?
  - How should we search and retrieve data of interest?
  - How should we get the best data in the right hands?
  - How should we then act on that data?
  - How do we do this quickly and seamlessly with as little investment as possible?
- **Platforms and Tools**
  - What infrastructure should I use?
  - What tools should I use?
  - Should I store on-premise or in the cloud?
  - Should I use open source?
  - How do I scale?

# Cleaning Data – Outlier Detection




**3-sigma outlier identifier on TD601 steam output data (moving window size=n/20=360)**

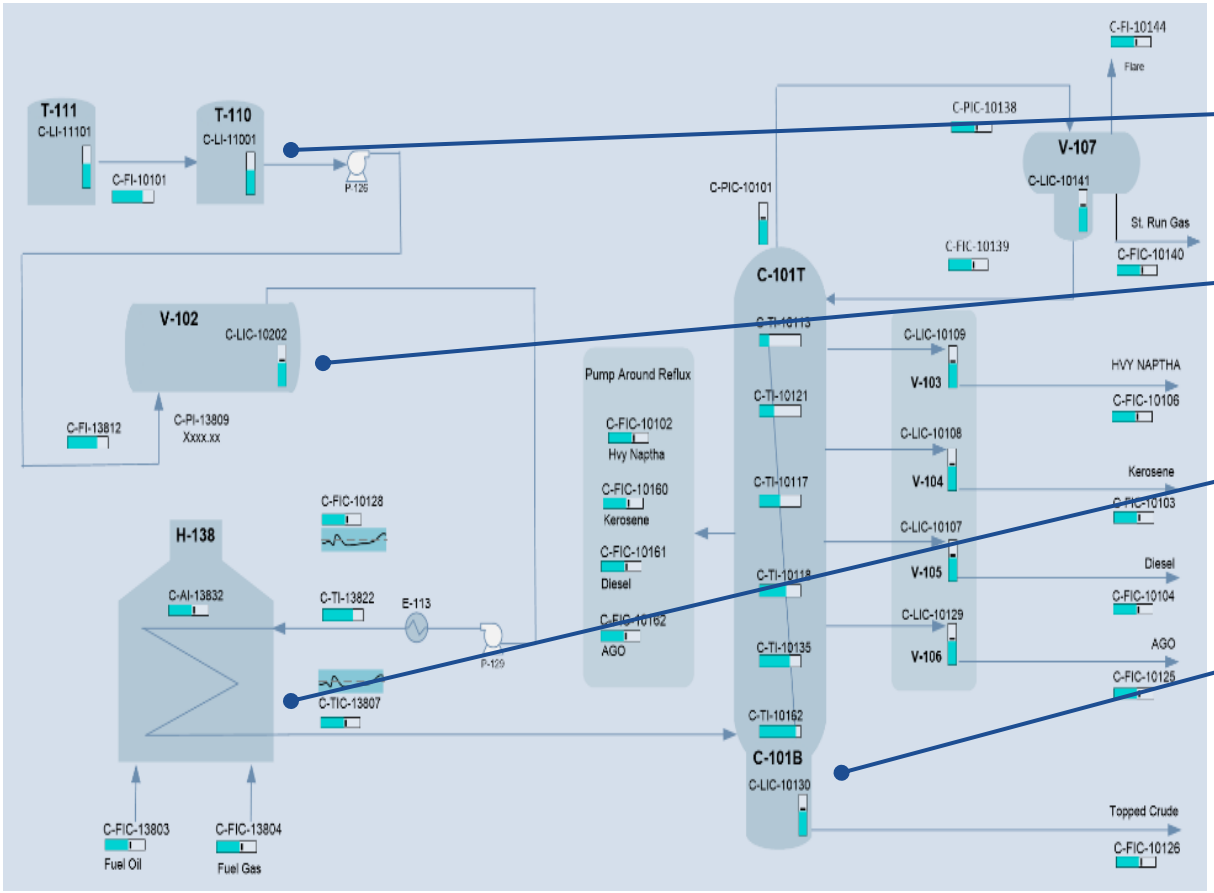



**TD601 steam output data processed by 3-sigma outlier identifier(moving window size=n/20=360)**

- **Compressed data**
- **Missing data**
- **Outliers**
- **Noisy data**

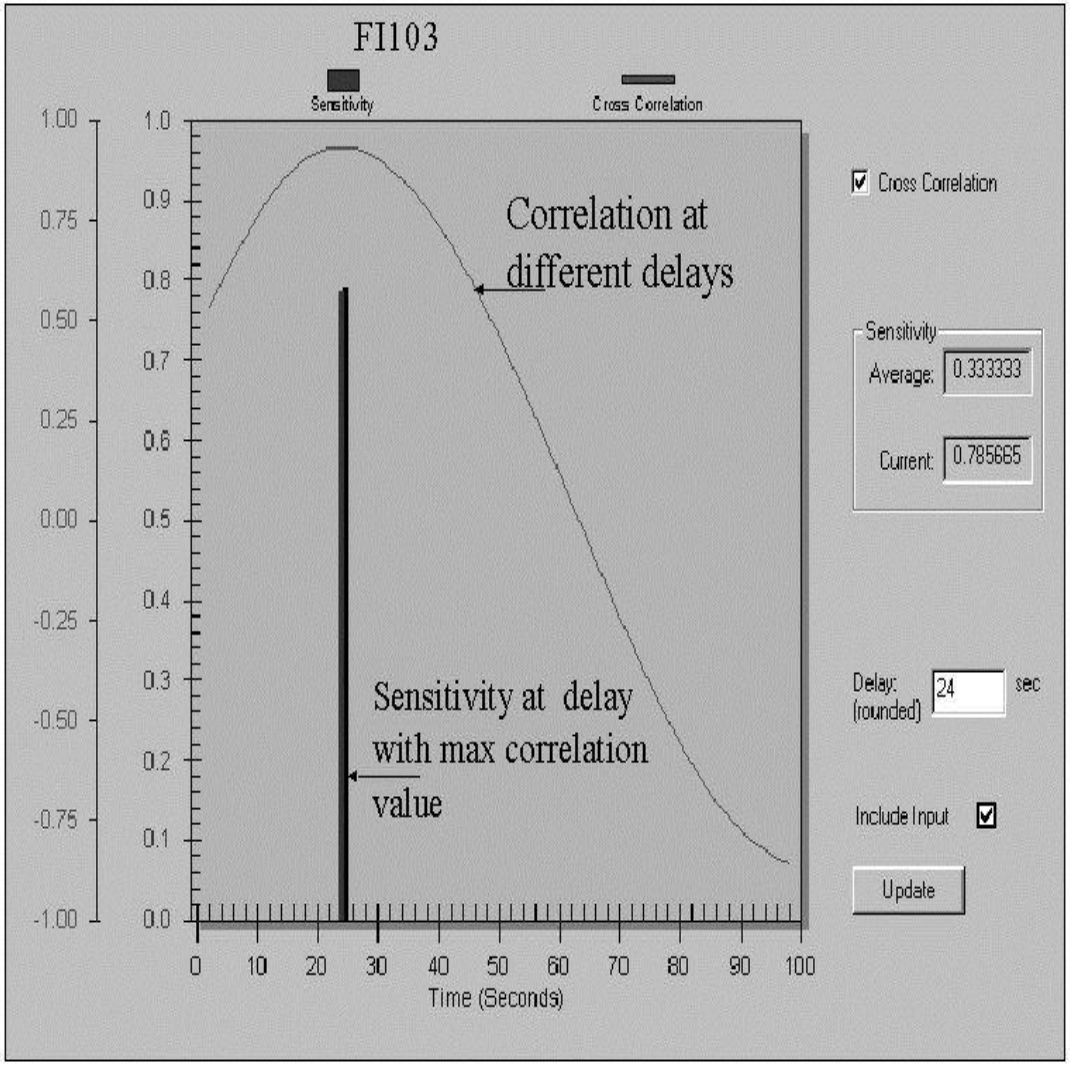
THE UNIVERSITY OF TEXAS AT AUSTIN
McKetta Department of Chemical Engineering
Cockrell School

# Time Relationships Matter



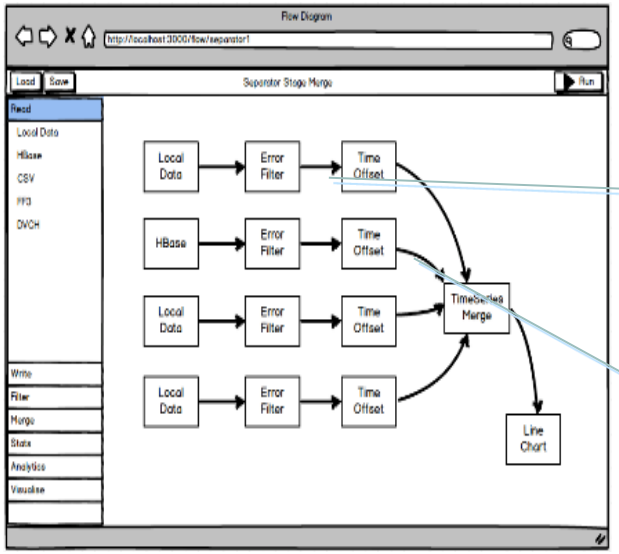

**Process Plant data analysis requires consideration of time delay of material transfer through process. In detecting relationships, individual variable time delays have to be identified along with the correlation/ causation impacts – problem dimensionality becomes much higher.**

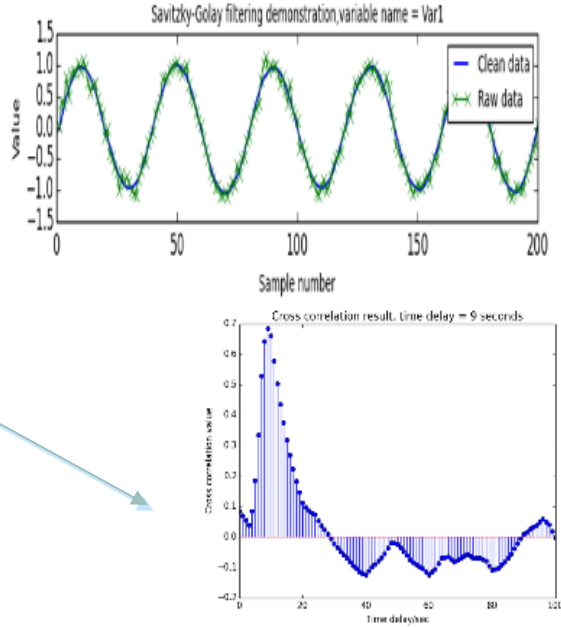
# Selecting Parameters And Time Delays



# Develop and Evaluate Models

**Utilize Block Diagramming Environment to develop and evaluate models**

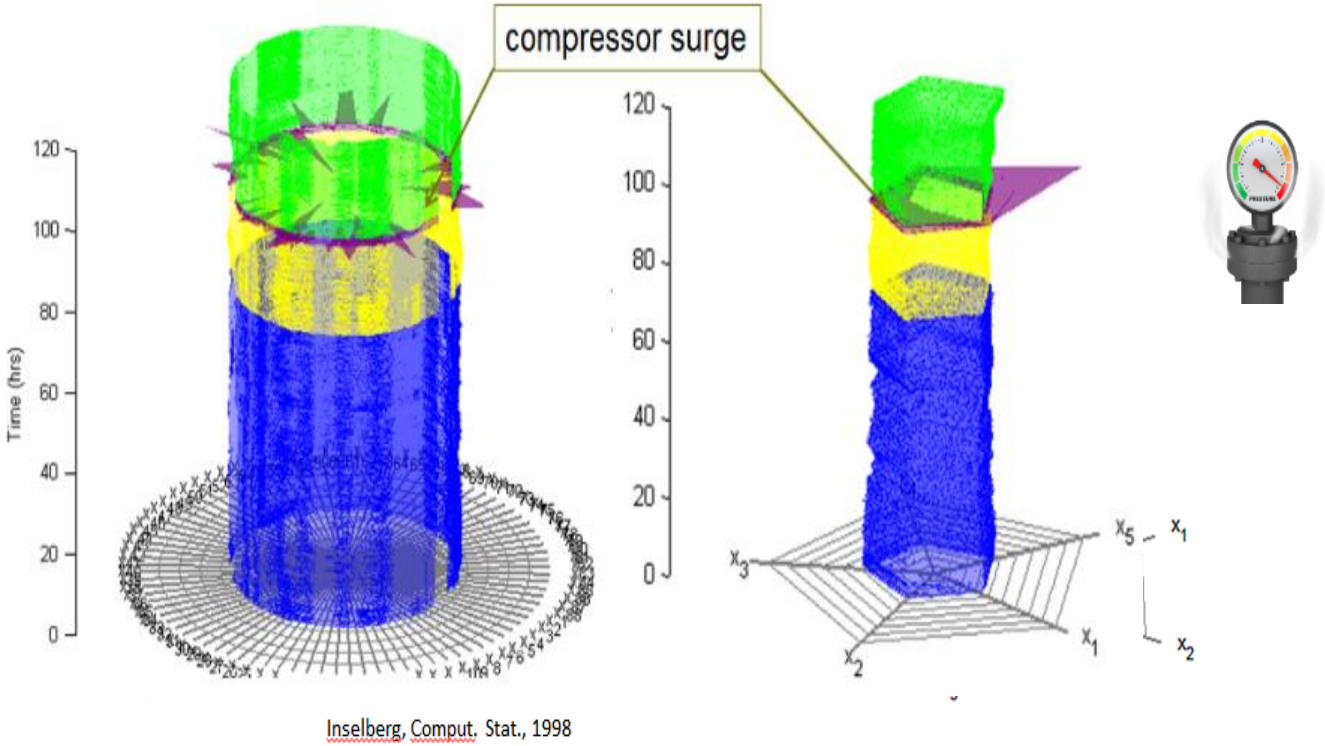



***Interact with your model visualizing results at each stage of the data processing, when you are done, deploy the model for run-time execution***

AIChE
Southwest Process Technology Conference

# Visualizing Predictions




Inselberg, Comput. Stat., 1998

**1 to 4 days before surge – Blue**
**1 day before surge – Yellow**
**2 hours before surge – Purple**
**Post surge - Green**

Center for Operator Performance
An Industry/University Center for Excellence

THE UNIVERSITY OF TEXAS AT AUSTIN
Cockrell School
McKetta Department of Chemical Engineering

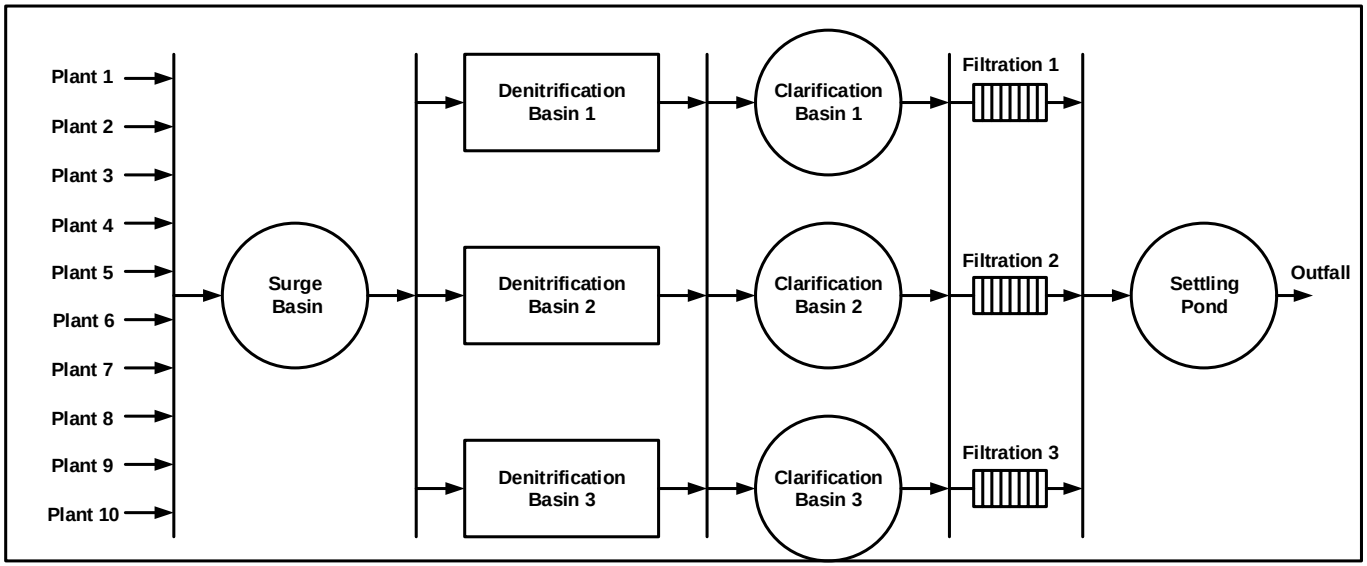
AIChE
Southwest Process Technology Conference


**Reference: Dunia, R; G. Rochelle; T.F. Edgar; M. Nixon; "Multivariate Modeling of a carbon dioxide removal process"; Computers & Chem Eng;60 (2014); pp. 381-395**

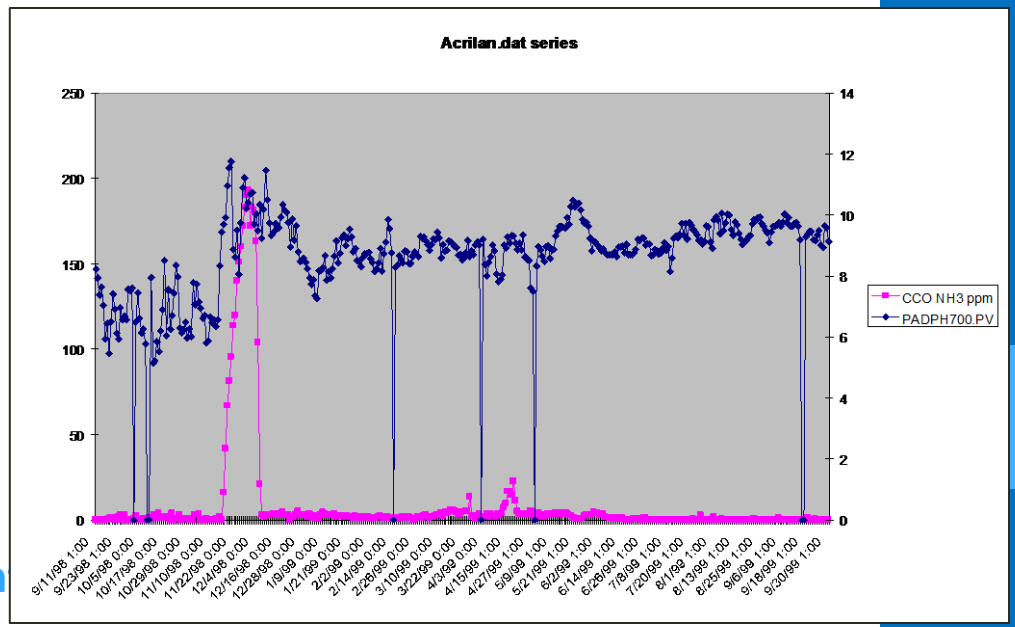
AFPM
American
Fuel & Petrochemical
Manufacturers

# Case Studies

# Case Study – Predictive Analytics – Waste Water Plant




- Multiple effluent NH3 spikes in outfall violating consent decree and incurring fines
- Residence time of the treatment plant >2 weeks making issue ID difficult
- More than 200 measured variables to be considered as possible predictors



# Case Study – Predictive Analytics – Waste Water Plant

- 5 years data analyzed for correlations with simultaneous identification of time lag
- Identification of strong correlation with pH of effluent from one of the plants – large drop in pH preceded excursion in denitrification basin by one day
- Vessel cleaning correspondence with pH drop

# Case Study – Predictive Analytics - Maintenance

- **Significance**
  - Sipchem Jubail site is considered a world-class complex with 4 large integrated plants
  - Successful implementation of complete integrated maintenance program
  - Built on installed base to create multi-phase asset management program
- **Project Scope**
  - Implementation of Asset Management including Reliability Centered Maintenance on 6,000 assets
  - Uses predictive analytics and diagnostics from field instrumentation
- **Audited Value**
  - 12% decrease in maintenance costs
  - 2% increase in Plant availability

Saudi International Petrochemical Company (Sipchem) actively develops and invests in petrochemical and chemical industries.

Sources: Abdullah Al Ghamde; John E. Hill,; "*Saudi International Petrochemical Company (Sipchem) Reliability Process,"* Paper 01A241; 2010 Emerson Global User Exchange; Sudheer Prabhakaran; John E. Hill,;"*Generating Actual Savings with Sipchem's Reliability Program,"* Paper 1B-2487; 2012 Emerson Global User Exchange

# Summary

- Predictive analytics is an evolving technology with many potential applications in the process industries
- Implementation of Predictive Analytics in the process industries has distinctive requirements/ issues due to the time series character and data
- Current applications in process fault detection, availability, safety and optimization have proven value – there will be more in the future

# Thank You For Attending. Questions?