

A young girl with dark hair, wearing a light-colored top and denim shorts, is sitting on the floor and playing chess with a boy. She is looking towards the camera with a slight smile. The chessboard and pieces are visible in the foreground. The background is a bright, out-of-focus indoor setting with large windows.

Brushing off Old Data

Dr Bob Sochon,
GSK Consumer Healthcare

What I'm talking about today



The last 58 years



1961

US drug company Block Drug launches Sensodyne Original, with strontium, the first desensitising toothpaste



1960s

Sensodyne Mint toothpaste launches with improved flavour



1970s

First formulations with potassium nitrate launch with a range of gentle toothbrushes



1970s

Sensodyne F, the first variant containing fluoride launches

1980s

Sensodyne launches Total Care F providing gum protection and sensitivity relief



2008

Sensodyne launches in India & China



2007

Sensodyne Iso-Active, the world's first gel-to-foam sensitivity toothpaste launches



2006

Sensodyne launches Pronamel, with an optimised fluoride formulation to help protect against acid erosion



2004

Sensodyne launches its first dentist testimonial TV advertising



2001

GSK acquires Block Drug and the Sensodyne business



2009

GSK acquires NovaMin®, a next generation active ingredient which helps repair sensitive teeth



2010

Sensodyne Rapid Relief launches offering relief in 60 seconds



2011

Sensodyne celebrates 50 years launching Repair & Protect with NovaMin® and fluoride, a toothpaste that can actually repair sensitive teeth



2013

Sensodyne Complete Protection launches, with 7 specially designed benefits in one sensitivity toothpaste



2014

Sensodyne True White launches, offering a breakthrough in whitening for sensitive teeth



2019

Sensodyne Sensitivity & Gum launches addressing 50% sufferers with both conditions



2018

Sensodyne wins Marketing Society of the Year Grand Prix and Euro Effies awards for advertising effectiveness.



2017

Sensodyne relaunches Rapid Relief, based on a superior stannous fluoride formulation



2016

Sensodyne becomes GSK's 1st £1 billion Consumer Healthcare Brand



2015

Sensodyne launches a mouthwash containing potassium, to protect sensitive teeth



Manufacturing Overview



Challenges & Opportunities

- Regulated industry
 - In many markets Sensodyne is regulated akin to a medicine
 - Change and experimentation is difficult
- Batch processes
 - Many isolated data sources
 - Never exactly the same day – day
 - Many interconnected steps



**Can we learn from our history to
make better products?**

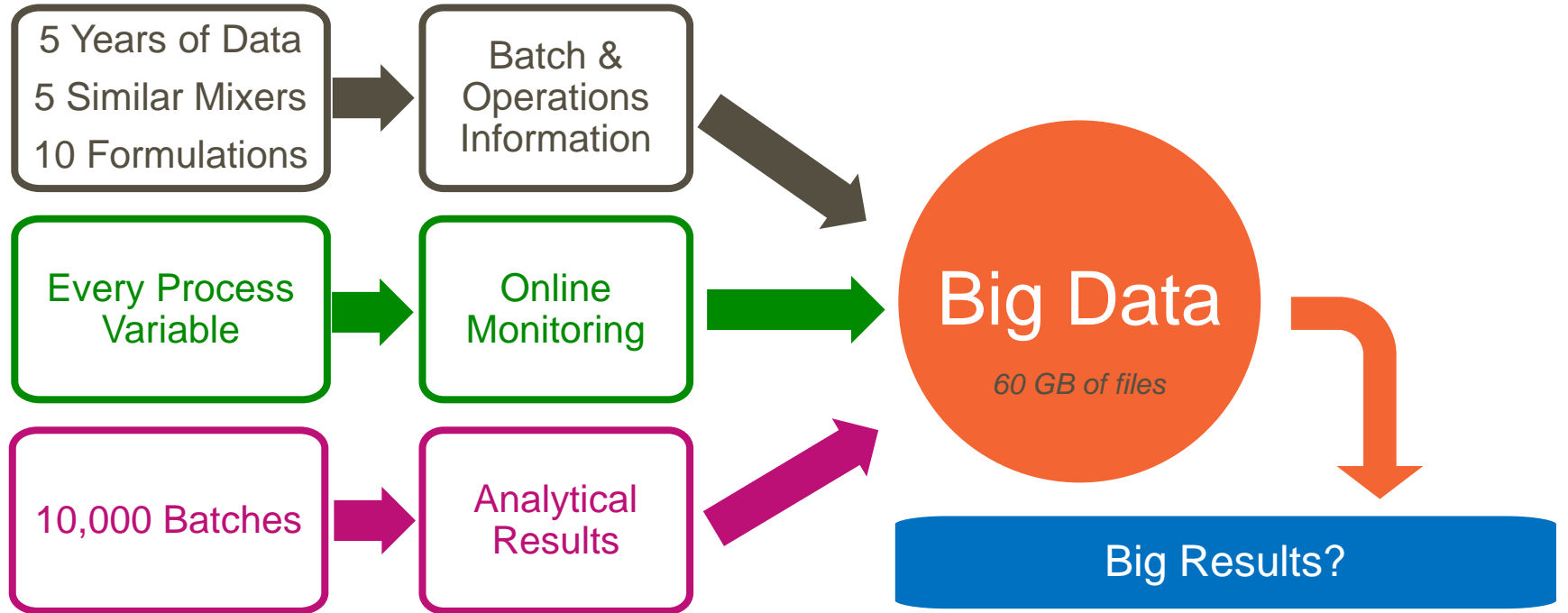


Big Data

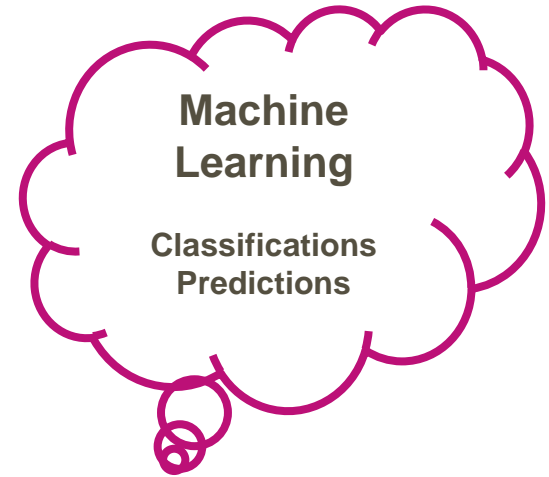
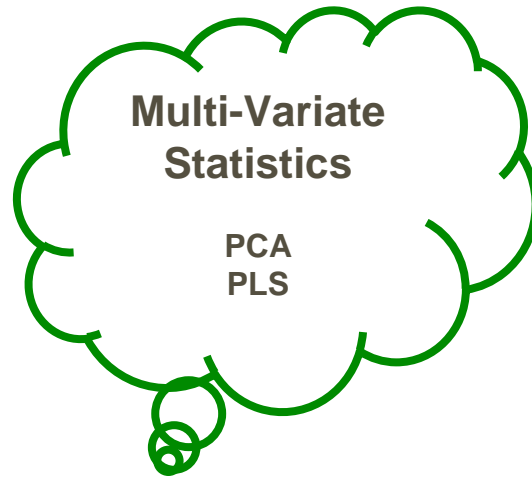
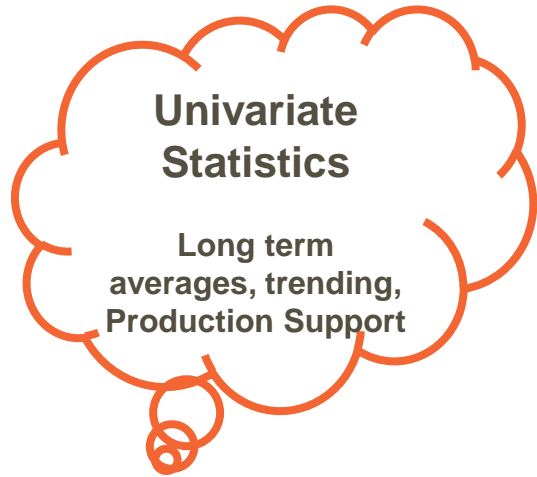
What does this mean?



What have we got?



What can we do with all this data?



Process Improvement and Understanding

Improved Manufacturing Efficiency & Reduced Cycle Time

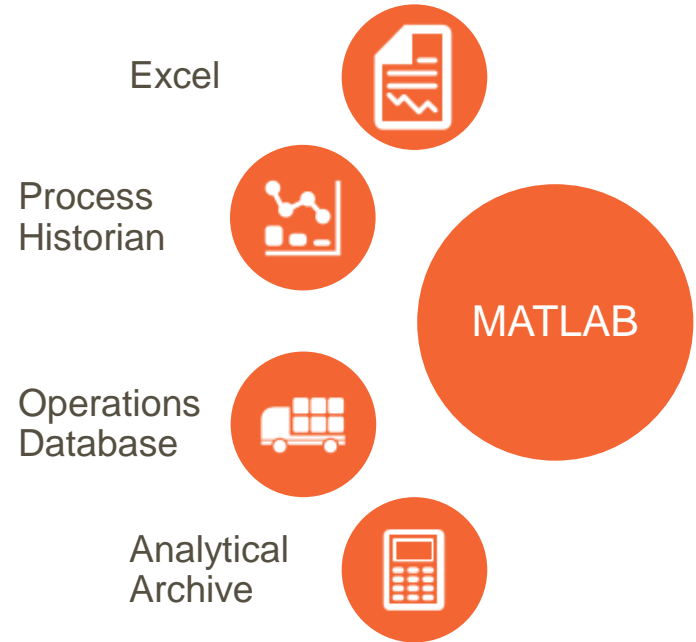
Improved R&D → Factory Product Transfer

Why MATLAB?



- Ability to link different data files & formats together
- Quickly iterate code
- Format / library independence
- Commercially supported / maintained
- Easy consulting support
- Build GUI to enable easy observation of data

Quick & easy code development cycles



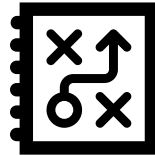


Results

What did we do?



Cleaning
& Context



Phase
Separation



Data
investigation



Results

Typical Workflow





Cleaning
& Context



Phase
Separation



Data
investigation



Results

Data Visualisation



Excel



Process Historian



Operations Database



Analytical Archive

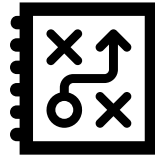


The screenshot shows a software interface with several panels. On the left, there are 'Data Folders' with fields for IP21 Data, Operator Data, LIMS Data, Phase Definitions, and Output Folder. Below that is a 'Mixer/Formulation Combination' table. The main area is titled 'Data Analysis' and 'Preprocessing', containing 'Run Stages' (Segment Batches, Segment Phases, Add LIMS data), 'Properties' (No. Batches, Skip Batches, Phase Cutoff, Verbose), and 'Batch Segmentation' (Stage, Start/End, Seg Var, Cut Direction, Cut Value, Window min/max, Window Midpoint, Smooth Range, Buffer Size). At the bottom, there are 'Off/On' and 'Safety Mode' controls.

Formula	Mixer Number
U3004	Mixer 1
U3008	Mixer 3
U3440	Mixer 4
U3442	Mixer 5
U3444	
U3530	
U3531	
U3532	



Cleaning
& Context



Phase
Separation

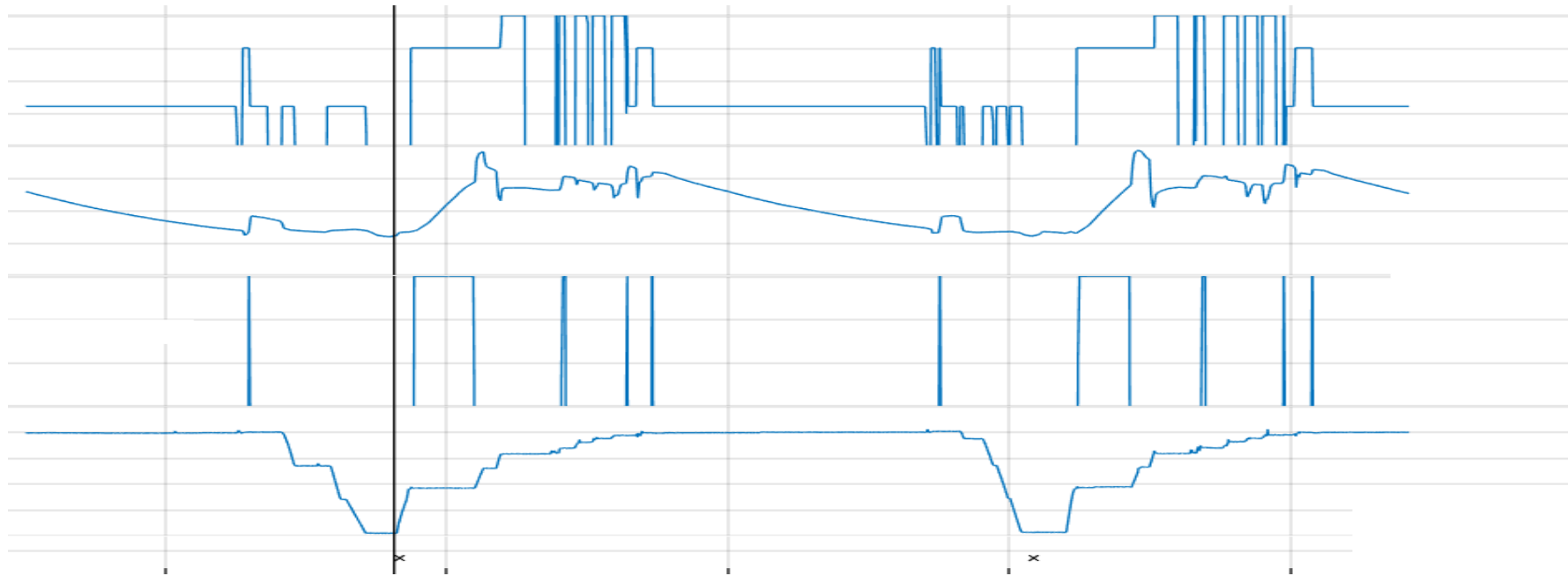


Data
investigation



Results

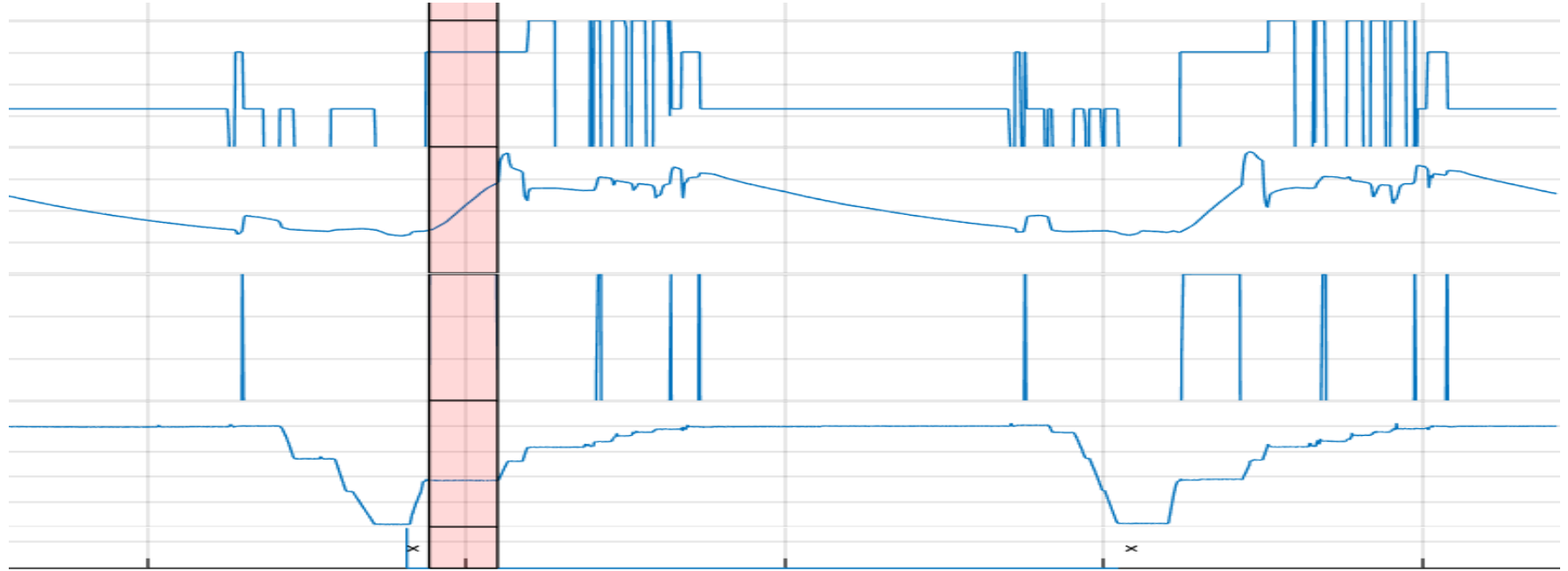
Typical Batch Trace



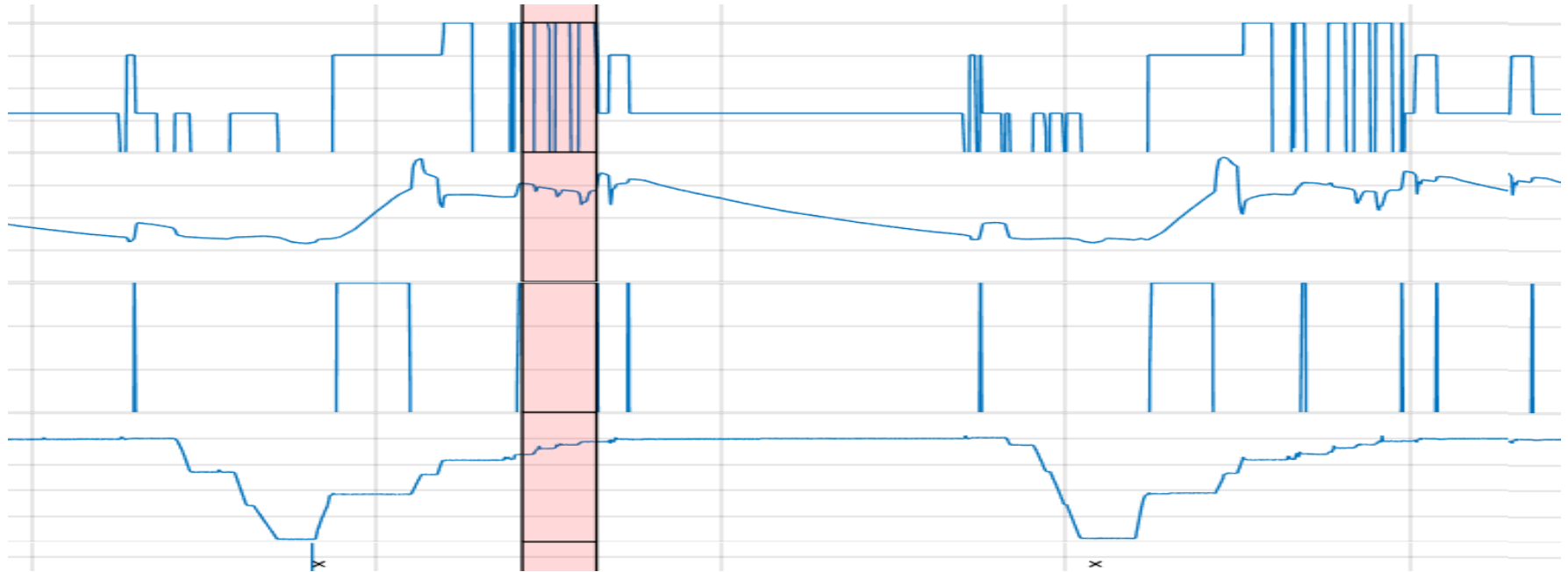
Overall Batch



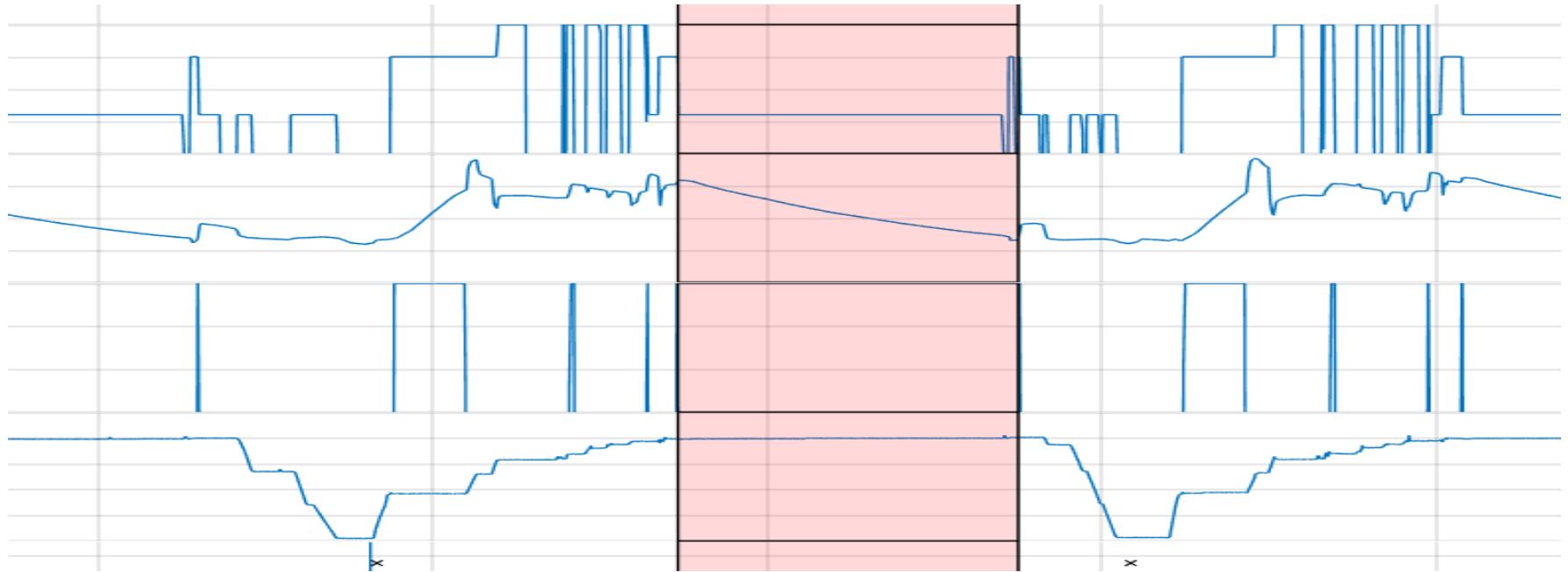
Batch Start-up



Adding Silica



Batch Finishing



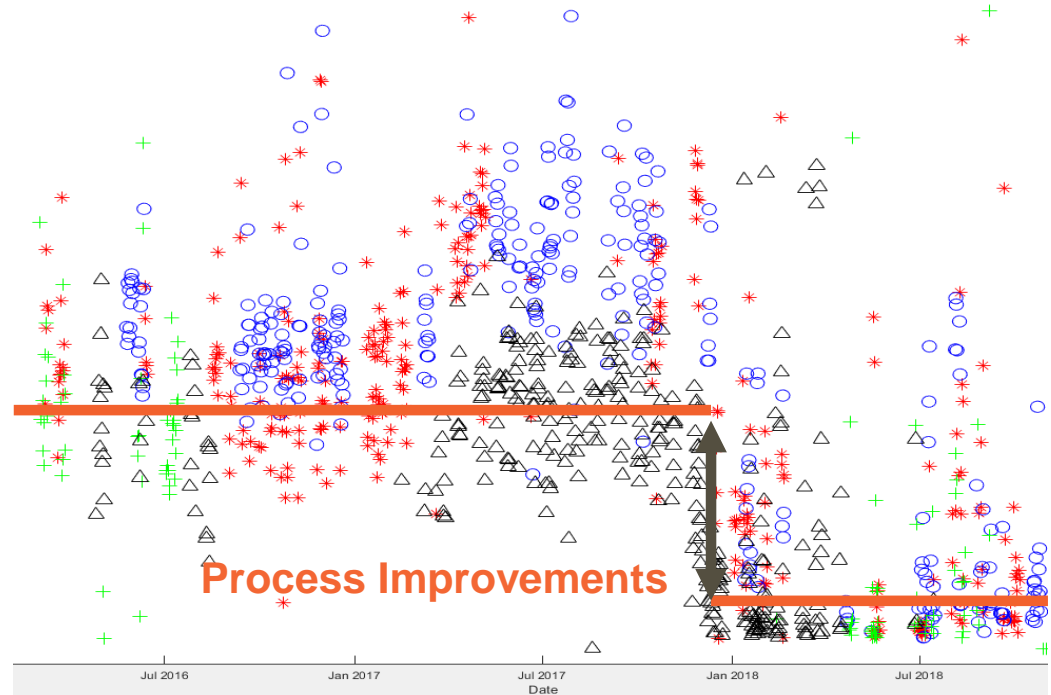
Improvements Over Time



Can now **automatically** pick phases out of data and start to ask simple questions:

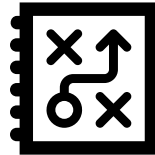
i.e:

- 1) Does process time improve over time?
- 2) Is processing time mixer dependant?
- 3) Do our process improvements work?





Cleaning
& Context



Phase
Separation

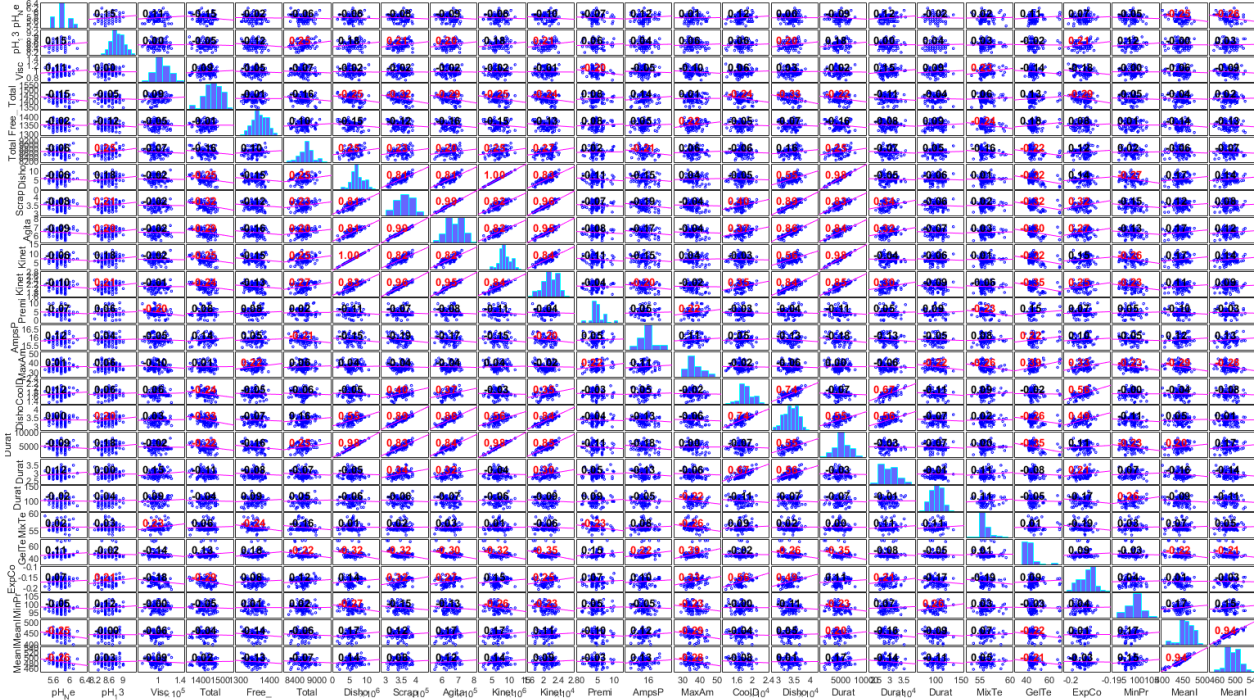


Data
investigation



Results

Linear Correlations

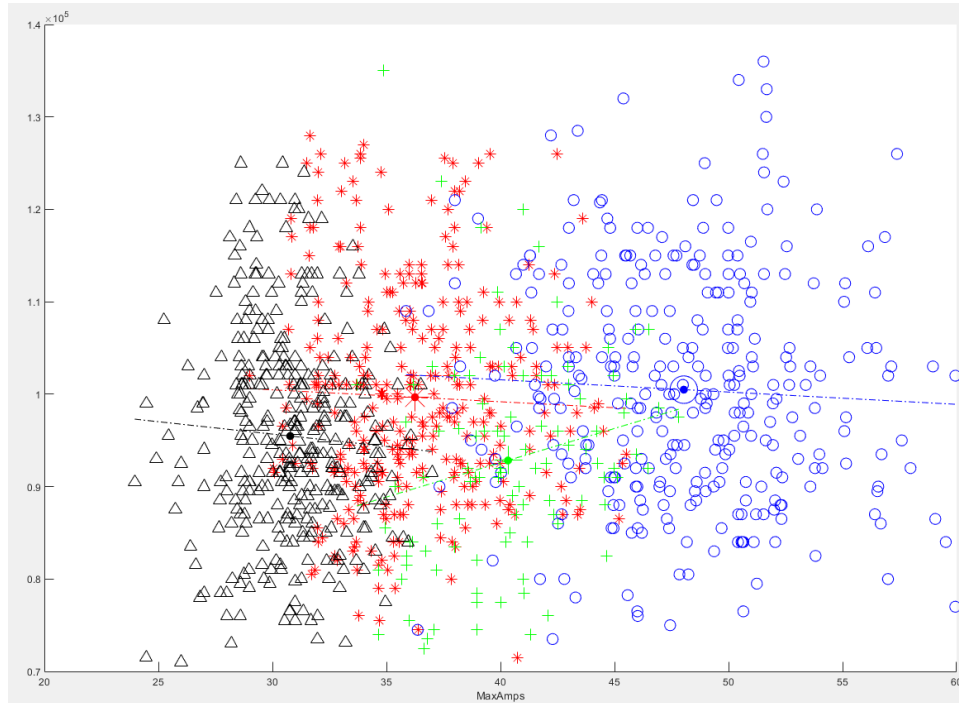


Too much Data to easily see!

Can see relationships in the data

Needs further understanding

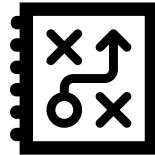
Linking Manufacturing Phases to Analytical Data



- Can link final batch properties to the batch to each phase of the data
- Not all mixers perform the same
 - Confirms our operator reports



Cleaning
& Context



Phase
Separation

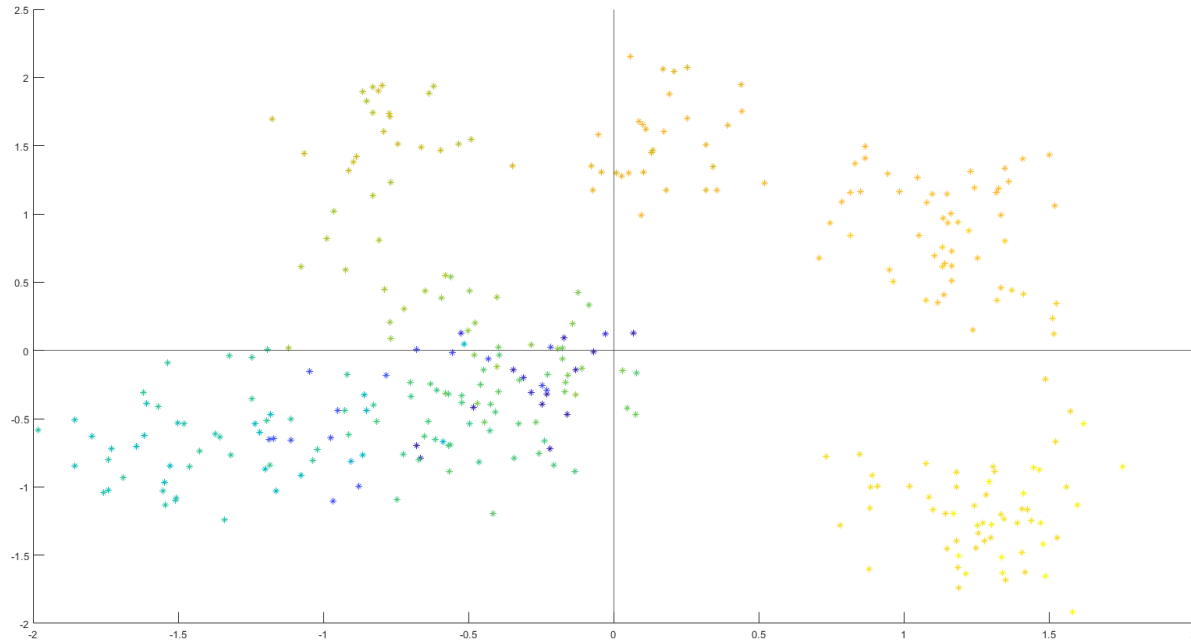


Data
investigation



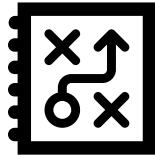
Results

Multivariate Understanding





Cleaning
& Context



Phase
Separation



Data
investigation



Results



What have we done?



-
- We had / have a lot of data that is very siloed, in many different systems
 - We finally have a way to match all Batch, Process & Analytical information together.
 - Can only get better as systems progresses in GSK
 - We can reliably deduce information about phasing and performance of the mixers, without human intervention.
 - Reduce time and error processing, easier conclusions.
 - Can do this across formulations, mixers (& years)
 - **We can now ask for conclusions in process data – i.e. does “this” effect “that”?**
 - Every growing as our factories make more batches!

What have we learnt?



-
- **Not all our variability is random**
 - For example different mixers seem more important than different processing times
 - More inputs will likely make better models, but never going to be perfect
 - **Big Data is not a panacea!** – It will not immediately solve all our problems
 - **We still have to do experiments to generate understanding**
 - Machine Learning / Multivariate Statistics can't replace science.
 - **Allows us to guide experiments to examine impacts of input variables.**



**Thank You For
Listening**

However: There are no SIMPLE easy correlations in the data

1. We might not have enough data:

Is there enough data here? – we only have 200 – 300 data points per batch

2. It might not vary enough to make clear versions of truth

Because of GMP / Factory nature of data i.e. mixer speeds seem mostly similar between batches

3. We might not be measuring the right thing

Our test methods don't exactly detect what we're looking for. Viscosity vs Rheology for example