PI World 2019 Lab

Advanced Analytics and Machine Learning Use Cases with Industrial Sensor Data

OSIsoft, LLC
1600 Alvarado Street
San Leandro, CA 94577 USA
Tel: (01) 510-297-5800
Web: http://www.osisoft.com

© 2019 by OSIsoft, LLC. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, mechanical, photocopying, recording, or otherwise, without the prior written permission of OSIsoft, LLC.

OSIsoft, the OSIsoft logo and logotype, Analytics, PI ProcessBook, PI DataLink, ProcessPoint, Asset Framework (AF), IT Monitor, MCN Health Monitor, PI System, PI ActiveView, PI ACE, PI AlarmView, PI BatchView, PI Vision, PI Data Services, Event Frames, PI Manual Logger, PI ProfileView, PI WebParts, ProTRAQ, RLINK, RtAnalytics, RtBaseline, RtPortal, RtPM, RtReports and RtWebParts are all trademarks of OSIsoft, LLC. All other trademarks or trade names used herein are the property of their respective owners.

U.S. GOVERNMENT RIGHTS

Use, duplication or disclosure by the U.S. Government is subject to restrictions set forth in the OSIsoft, LLC license agreement and as provided in DFARS 227.7202, DFARS 252.227-7013, FAR 12.212, FAR 52.227, as applicable. OSIsoft, LLC.

Published: April 24, 2019

Data Science – Advanced Analytics and Machine Learning Use Cases – Hands-on Lab – OSIsoft PI World 2019

Lead: Gopal GopalKrishnan, P.E., Solution Architect, Partners & Strategic Alliances

Lead: Curt Hertler, Principal Technical Advisor, Delivery Services

Instructor: Alassane Seck, Systems Engineer

Table of Contents

Table of Contents 3
Advanced Analytics and Machine Learning Use Cases with Industrial Sensor Data 5
Lab Description 5
Summary 5
Exercise 1: Monitoring the batch evolution 6
Exercise 2: Full batch assessment 8
Exercise 3: Predict quality 8
Installed Software 10
PI System 10
Part I – Explore Yeast Manufacturing Operations and Extract Data using PI Integrator 11
Process Description 11
Step 1: Explore the AF model 12
Step 2: Explore yeast batch Event Frames 13
Data Publication 16
Step 1: Training dataset – Batch Summary 16
Select Data 16
Modify View 18
Publish 19
Step 2: Training dataset – Batch Evolution 19
Select Data 19
Modify View 21
Publish 21
Step 3: Test dataset – Batch Summary 22
Step 4: Test dataset – Batch Evolution 22
Part II – Model Development Using R 23
Exercise 1 – Monitoring the Batch Evolution 23
Part II - Model deployment and real-time scoring 25
Step 1: AF Analytics 25
Step 2: (Optional) Copy/paste the equations from R to AF Analytics 25
Step 3: Real-time scoring 26
Configure PI Notifications to receive an email (optional) 26
Replay the data for Batch U3054 28
Part III – Model Development Using R – Full Batch Assessment 31

Exercise 2 Full Batch Assessment 31
Part IV – Model Development Using R – Predict Quality 31
Exercise 3 Predict Quality 31
Reference Materials 32

Advanced Analytics and Machine Learning Use Cases with Industrial Sensor Data

Lab Description

In previous years, we have explored the use of advanced analytics and machine learning for:

>Anomaly detection in an HVAC air-handler - more

>RUL (remaining useful life) prediction based on engine operations and failure data - more

>Golden-run identification for the temperature profile from a feed dryer (silica gel/molecular sieve) in an oil refinery - more

Additionally, as part of the above labs, we have used analytical methods such as PCA (principal components), SVM (support vector), shape similarity measures etc. And, in other similar labs, we have covered well-known algorithms for regression, classification etc. and reviewed the use of Azure Machine Learning - more - and open source platforms such as R and Python.

In this year's lab, we explore the use of historical process data to predict quality and yield for a product (Yeast) in batch manufacturing. We'll use multivariate PCA modeling to walk-through the diagnostics for monitoring the 14-hour evolution of each batch. And, alert you when a batch may go "bad" as critical operating parameters violate "golden batch" criteria ((high pH, low Molasses etc.). And, then we utilize PLS – projection to latent structures - to predict product quality and yield at batch completion.

The lab illustrates the end-to-end tasks in a typical data science project – from data preparation, conditioning, cleansing etc. to model development using training data, testing/validation using unseen data, and finally, deployment for production use with real-time data.

The techniques explored in the lab are not limited to batch manufacturing; they can be applied to several industries and to numerous processes that are multivariate in nature.

No coding or prior experience with open source R or Python is necessary but familiarity with the PI System is a pre-requisite.

Who should attend? Power User and Intermediate

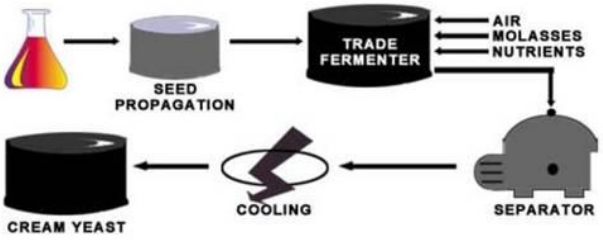
Duration: 3 hours

All products, names, trademarks and registered trademarks are property of the respective owners. All company, product and service names used in this document are for identification purposes only. Use of these names, trademarks, brands, and any OSIsoft partner software during the hands-on lab does not imply endorsement.

Summary

In this lab, we review Yeast manufacturing operations – specifically, the fermentation process. A typical batch of yeast fermentation takes 13 to 14 hours. There are many variables that can affect the yield and

quality of a batch: feed variability (molasses, air, NH3), byproducts (ethanol), bioreactor conditions (temperature, pH). Any combination of these factors can result in “bad” batch runs.



We want to use historic operations data with known “good” runs as a basis for alerts when current production parameters deviate from “golden batch” conditions. We also want to predict quality parameters, generically referred as QP1 and QP2, and the expected yield for each batch.

Bioreactor

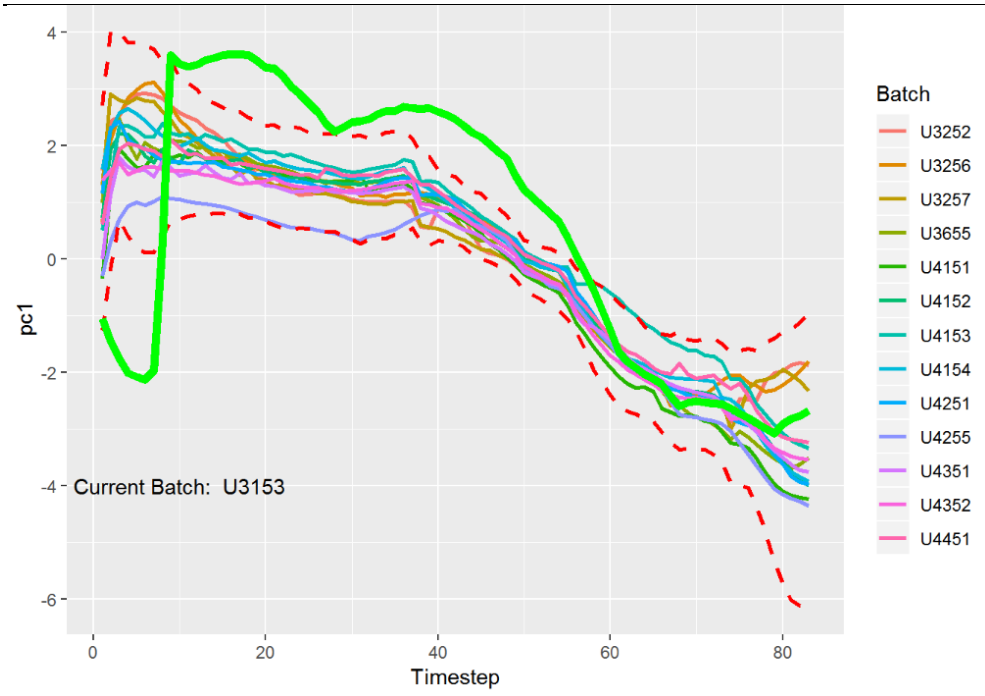
General | Child Elements | Attributes | Ports | Analyses | Notification Rules | Version

Filter

Name	Value	Time Stamp	Description
Category: Batch Info			
Batch	None	3/5/2019 12:00:00 AM	Batch ID
Batch Active	No	3/5/2019 12:00:00 AM	Active Fermentation Indicator
Status	Good	3/4/2019 11:30:00 PM	Batch Status : Manual Entry from Operator
Category: Lab Data			
Amount	5977 kg	3/4/2019 11:30:00 PM	Mass of yeast in bioreactor
Innoc	973 g	3/4/2019 11:30:00 PM	Mass of Yeast in Innoculation
QP1	91	3/4/2019 11:30:00 PM	Quality Parameter 1
QP2	80	3/4/2019 11:30:00 PM	Quality Parameter 2
Yield	0.5 %	3/4/2019 11:30:00 PM	Yield of fermentation
Category: Process Data			
Air	0 m3/h	3/5/2019 12:00:00 AM	Air Flowrate
Ethanol	0 mg/L	3/5/2019 12:00:00 AM	Ethanol Concentration
Level	0 %	3/5/2019 12:00:00 AM	Bioreactor Level
Molasses	0 m3/h	3/5/2019 12:00:00 AM	Molasses Flowrate
NH3	0 L/h	3/5/2019 12:00:00 AM	NH3 Flowrate
pH	7	3/5/2019 12:00:00 AM	Bioreactor pH
Temp	25 °C	3/5/2019 12:00:00 AM	Bioreactor Temperature

Exercise 1: Monitoring the batch evolution

In this Exercise, we will use PCA model to develop the multivariate metrics to monitor the evolution of the batch as it is running and alert when conditions deviate from the model defined 3-sigma limits.



Add a new variable

Evaluate

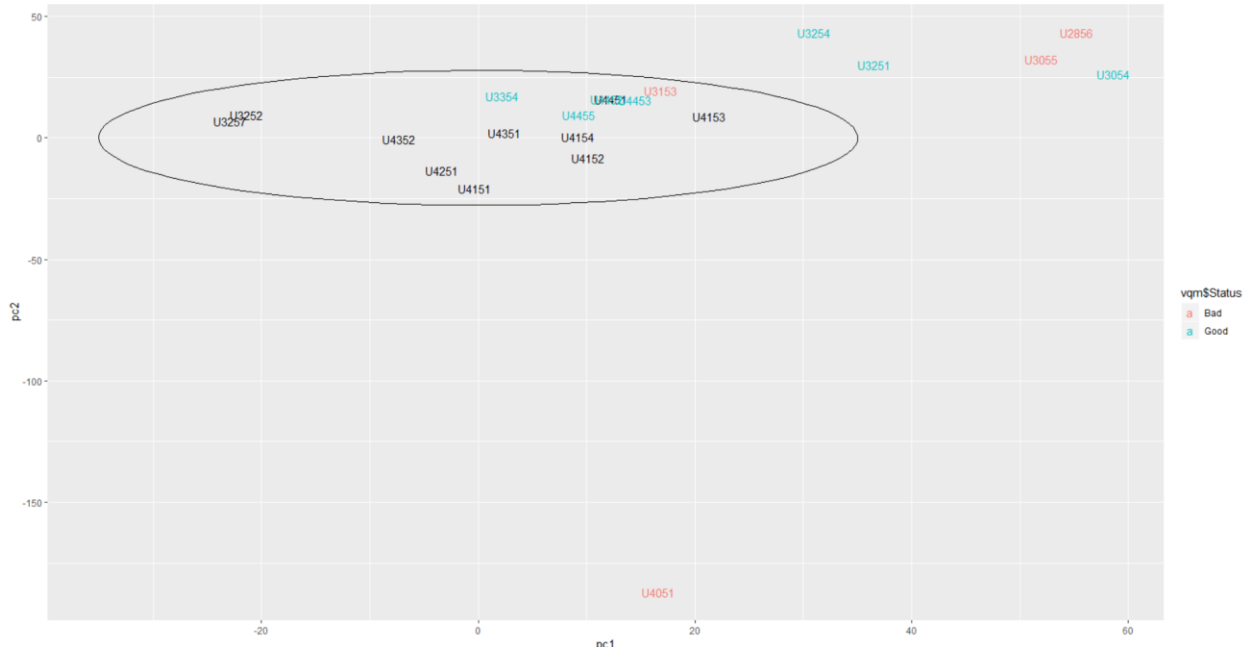
Name	Expression	Output Attribute
BatchAct	'Batch Active' = 1	Map
PC1var	If BatchAct Then ('Ethanol'-(0.6546))/0.4914*0.3986+('Temp'-(31.16))/1.226*-0.4698+('Molasses'-(2136))/95	PC1

```
If BatchAct Then ('Ethanol'-(0.6546))/0.4914*0.3986+('Temp'-(31.16))/1.226*-0.4698+('Molasses'-
(2136))/957.1*-0.06959+('NH3'-(107.5))/63.6*0.336+('Air'-(5872))/1444*-0.3142+('Level'-(48.63))/5.067*-0.5222+('pH'-
(5.205))/0.3461*-0.3622 Else DigState("No Data")
```

PC2var	If BatchAct Then ('Ethanol'-(0.6546))/0.4914*-0.1048+('Temp'-(31.16))/1.226*0.1278+('Molasses'-(2136))/95	PC2
T2var	//Hotelling T2 If BatchAct Then (PC1var/1.836)^2+(PC2var/1.402)^2 Else DigState("No Data")	T2

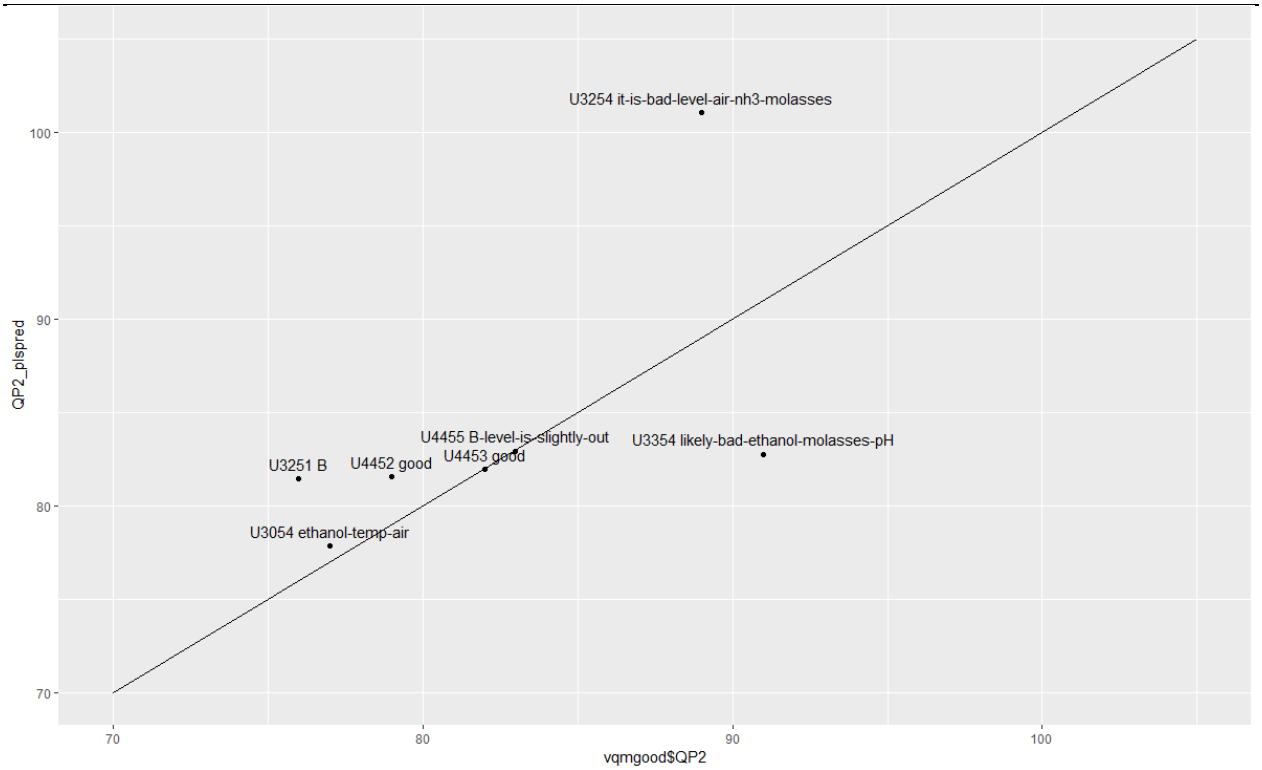
Exercise 2: Full batch assessment

In this Exercise, we will use the full batch run data – with PCA to make an assessment regarding "good" vs. "bad" batch. This is useful when the results from lab quality analysis may not be available for 4-8 hours after the end of a batch.



Exercise 3: Predict quality

In the portion of the lab, we will use PLS to predict quality.


U3254 it-is-bad-level-air-nh3-molasses
U4455 B-level-is-slightly-out
U3354 likely-bad-ethanol-molasses-pH
U4453 good
U3251 B
U4452 good
U3054 ethanol-temp-air
QP2_plspred
vqmgood$QP2
100
90
80
70
70
80
90
100

Installed Software

PI System

The VM (virtual machine) used for this lab has the following PI System software installed:

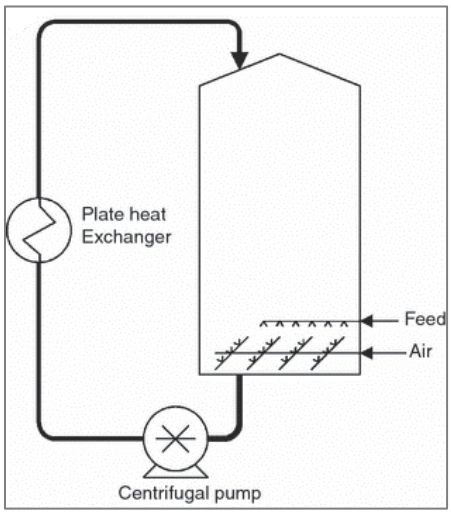
Software	Version
PI Data Archive	2018
PI Asset Framework (PI AF) server	2018
PI Asset Framework (PI AF) client (PI System Explorer)	2018
PI Analysis & PI Notifications Services	2018
PI Vision	2017 R2 SP1
PI Web API	2018
PI Integrator for Business Analytics	2018 R2

For details on PI System software, please refer to:
http://www.osisoft.com/pi-system/pi-capabilities/product-list/

Part I – Explore Yeast Manufacturing Operations and Extract Data using PI Integrator

Process Description

We have a plant manufacturing Baker's yeast in a fed-batch bioreactor with a circulation loop, shown below.



- Air is supplied via an air sparger, and mixing ensures tank aeration.
- Molasses is fed using perforated pipes at the bottom of the bioreactor.
- Ethanol concentration is monitored using an in-line sensor.
- Acidity in the bioreactor is monitored using a pH sensor in the circulation loop.
- A plate heat exchanger is used to control temperature, which is monitored using a thermocouple.
- Bioreactor level is also monitored via a sensor.
- All process signals described above are transmitted to a programmable logic controller (PLC)
- The operator enters the Batch ID, and whether the batch is "Active" in the PLC through a Human Machine Interface (HMI). They can also manually enter if the batch has been "Good" or "Bad" at the end of the batch.

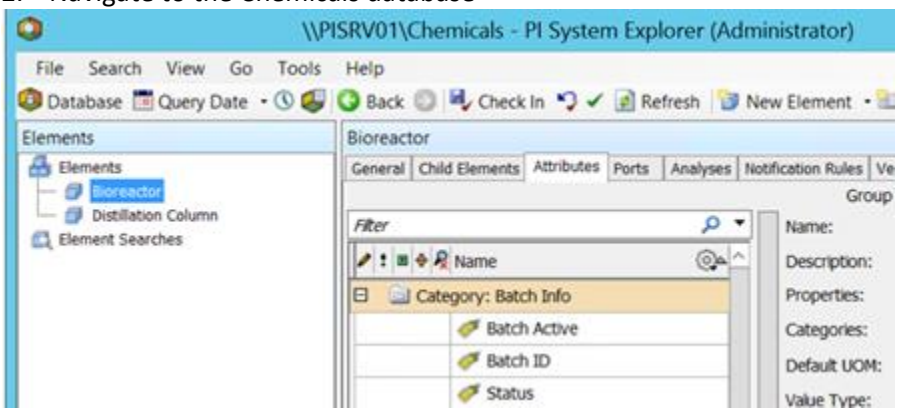
At the beginning of the batch, a sample of the inoculum is sent to the lab. When the batch is completed, samples of the bioreactor broth are taken for analysis. Lab data is measured and/or calculated for each batch, and entered in the laboratory information management system (LIMS). This includes:

- The mass of yeast in the inoculum
- The final amount of yeast in the bioreactor
- The yield
- Two additional quality parameters, referred generically as QP1 and QP2

Step 1: Explore the AF model

An AF model has already been created for the yeast production process.

1. Run PI System Explorer
2. Navigate to the Chemicals database



3. Select “Bioreactor”, and navigate to the “Attributes” tab.

Bioreactor

General | Child Elements | Attributes | Ports | Analyses | Notification Rules | Version

Filter

Name	Value	Time Stamp	Description
Category: Batch Info			
Batch	None	3/5/2019 12:00:00 AM	Batch ID
Batch Active	No	3/5/2019 12:00:00 AM	Active Fermentation Indicator
Status	Good	3/4/2019 11:30:00 PM	Batch Status : Manual Entry from Operator
Category: Lab Data			
Amount	5977 kg	3/4/2019 11:30:00 PM	Mass of yeast in bioreactor
Innoc	973 g	3/4/2019 11:30:00 PM	Mass of Yeast in Innoculation
QP1	91	3/4/2019 11:30:00 PM	Quality Parameter 1
QP2	80	3/4/2019 11:30:00 PM	Quality Parameter 2
Yield	0.5 %	3/4/2019 11:30:00 PM	Yield of fermentation
Category: Process Data			
Air	0 m3/h	3/5/2019 12:00:00 AM	Air Flowrate
Ethanol	0 mg/L	3/5/2019 12:00:00 AM	Ethanol Concentration
Level	0 %	3/5/2019 12:00:00 AM	Bioreactor Level
Molasses	0 m3/h	3/5/2019 12:00:00 AM	Molasses Flowrate
NH3	0 L/h	3/5/2019 12:00:00 AM	NH3 Flowrate
pH	7	3/5/2019 12:00:00 AM	Bioreactor pH
Temp	25 °C	3/5/2019 12:00:00 AM	Bioreactor Temperature

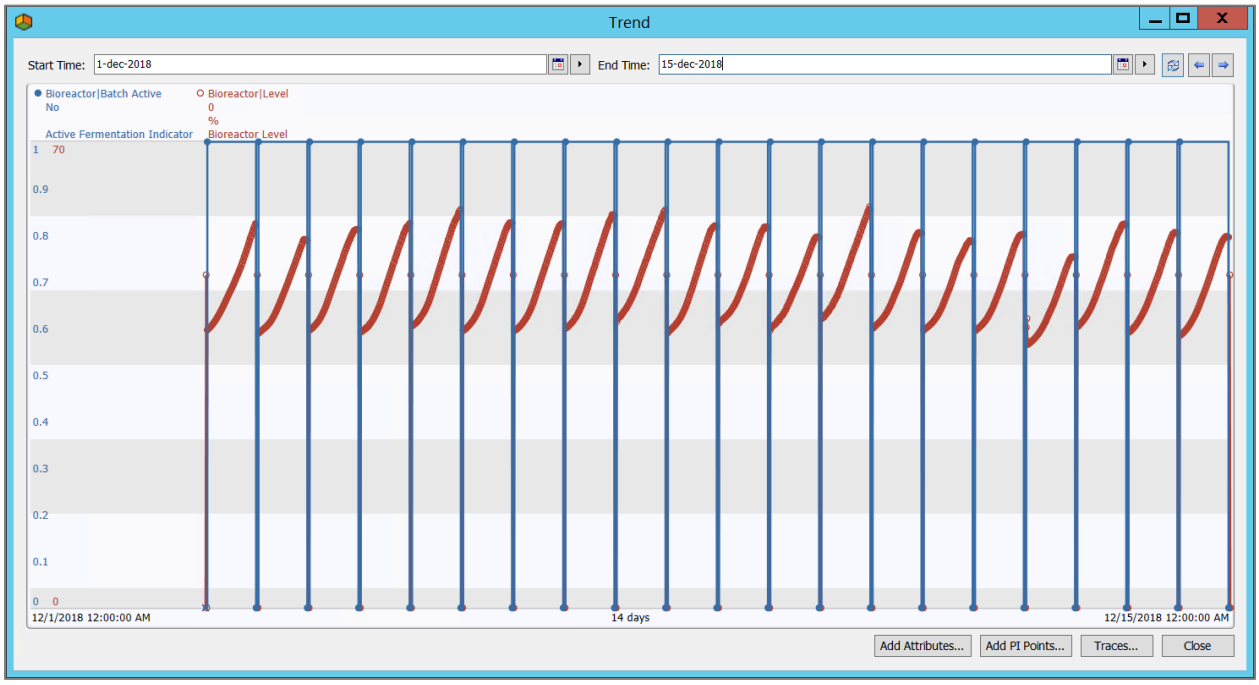
The attributes under “Batch Info” and “Process Data” are collected in near real-time. The “Lab Data” is collected by the LIMS where the lab sample uses the batch end timestamp for the lab results.

Step 2: Explore yeast batch Event Frames

We use asset analytics to generate the Event Frames needed to analyze yeast batch data.

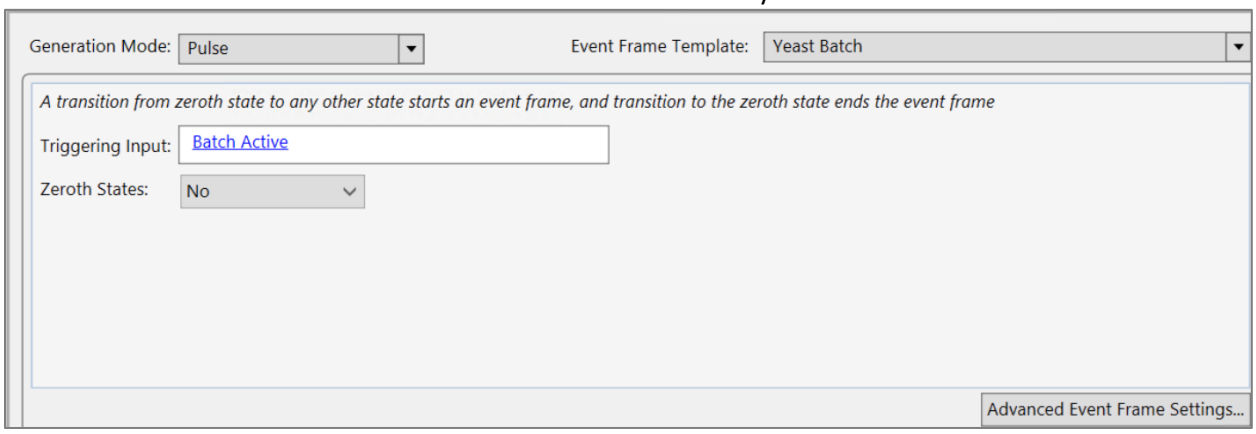
1. In the Bioreactor element, navigate to "Attributes" tab
2. While holding the CTRL key, select the "Batch Active" and "Level" attributes
3. Right-click and select "Trend"
4. Change the start time to 1-dec-2018 and the end time to 15-dec-2018

You should see the following trend:



In the trend, the "Batch Active" value is "Yes", an active batch is running and hence it is used to generate the Event Frames.

1. Close the trend window
2. Navigate to the "Analyses" tab
3. Review the "Yeast Batch" Event Frame Generation Analysis.



In the “Pulse” generation mode, by setting the Triggering Input to “Batch Active”, and Event Frame will be created when the value changes to “Yes”, and will close when the value changes to “No”.

To see the end result:

1. Navigate to the Event Frames tab
2. Right-click on “Event Frame Searches” and select “New Search”
3. In the Event Frame Search Window, select the Template “Yeast Batch” and click OK

You will see several Event Frames:

Yeast Batches

Filter

Name	Duration	Start Time	Batch	Status	Amount
U2952 - Yeast Batch	13:40:00	12/3/2018 1:30:00 AM	U2952	No Data	No Data
U2953 - Yeast Batch	13:40:00	12/3/2018 3:50:00 PM	U2953	No Data	No Data
U3152 - Yeast Batch	13:40:00	12/4/2018 6:10:00 AM	U3152	Bad	6368 kg
U3252 - Yeast Batch	13:40:00	12/4/2018 8:30:00 PM	U3252	Good	6658 kg
U3253 - Yeast Batch	13:40:00	12/5/2018 10:50:00 AM	U3253	No Data	No Data
U3256 - Yeast Batch	13:40:00	12/6/2018 1:10:00 AM	U3256	Good	7120 kg
U3257 - Yeast Batch	13:40:00	12/6/2018 3:30:00 PM	U3257	Good	6920 kg
U3351 - Yeast Batch	13:40:00	12/7/2018 5:50:00 AM	U3351	No Data	No Data
U3451 - Yeast Batch	13:40:00	12/7/2018 8:10:00 PM	U3451	No Data	No Data
U3655 - Yeast Batch	13:40:00	12/8/2018 10:30:00 AM	U3655	Good	6597 kg
U4151 - Yeast Batch	13:40:00	12/9/2018 12:50:00 AM	U4151	Good	5541 kg
U4251 - Yeast Batch	13:40:00	12/9/2018 3:10:00 PM	U4251	Good	5749 kg
U4255 - Yeast Batch	13:40:00	12/10/2018 5:30:00 AM	U4255	Good	5365 kg
U4152 - Yeast Batch	13:40:00	12/10/2018 7:50:00 PM	U4152	Good	5875 kg
U4153 - Yeast Batch	13:40:00	12/11/2018 10:10:00 AM	U4153	Good	5835 kg
U4154 - Yeast Batch	13:40:00	12/12/2018 12:30:00 AM	U4154	Good	5973 kg
U4155 - Yeast Batch	13:40:00	12/12/2018 2:50:00 PM	U4155	Bad	5402 kg
U4351 - Yeast Batch	13:40:00	12/13/2018 5:10:00 AM	U4351	Good	5982 kg
U4352 - Yeast Batch	13:40:00	12/13/2018 7:30:00 PM	U4352	Good	5977 kg
U4451 - Yeast Batch	13:40:00	12/14/2018 9:50:00 AM	U4451	Good	5977 kg
U2856 - Yeast Batch	13:40:00	2/28/2019 5:50:00 AM	U2856	Bad	5715 kg
U2954 - Yeast Batch	13:40:00	2/28/2019 8:10:00 PM	U2954	No Data	No Data
U3054 - Yeast Batch	13:40:00	3/1/2019 10:30:00 AM	U3054	Good	6089 kg
U3055 - Yeast Batch	13:40:00	3/2/2019 12:50:00 AM	U3055	Good	5904 kg

Double-click on an Event Frame, and navigate to the “Attributes” tab.

U4255 - Yeast Batch

General | Child Event Frames | Referenced Elements | Attributes

Filter

Name	Value	Time Stamp
Category: Batch Info		
Batch	U4255	2/28/2019 5:30:00 AM
Status	Good	2/28/2019 7:10:00 PM
Category: Lab Data		
Amount	5365 kg	2/28/2019 7:10:00 PM
Innoc	914 g	2/28/2019 7:10:00 PM
QP1	93	2/28/2019 7:10:00 PM
QP2	77	2/28/2019 7:10:00 PM
Yield	0.48 %	2/28/2019 7:10:00 PM
Category: Process Data		
Max Ethanol	0.85934 mg/L	2/28/2019 12:30:00 PM
Max pH	6.9551	2/28/2019 7:10:00 PM
Min pH	5.1592	2/28/2019 7:00:00 AM
Total Air	80328 m3	2/28/2019 7:10:00 PM
Total Molasses	33426 m3	2/28/2019 7:10:00 PM
Total NH3	1622.3 L	2/28/2019 7:10:00 PM

The attribute configuration in the “Yeast Batch” Event Frame template (found under Library) contains the lab data, and other process data aggregates.

Note that some batches have “No Data” under the Lab Data attributes. We may filter out these batches during analysis.

Data Publication

Now that our batch data has been properly contextualized using Event Frames, we need to publish the historical data in a format suitable for multivariate modelling in R/Python.

Using PI Integrator for Business Analytics., we create two datasets:

- Training - data from **1 – 15 December 2018** for training and developing a model.
- Test - data from **28 February – 8 March** to test/validate the model.

For each dataset, we publish two tables:

- **Batch Summary table**: a table with initial conditions (inoculum) and final properties (quality, amount, yield) for each batch
- **Batch Evolution table**: a table that shows the evolution of the process data for the bioreactor (molasses, air, temperature, pH, ethanol). We will publish interpolated data at 10 minute intervals.

At the end of this step, you see 4 publications:

1. Training dataset – Batch Summary
2. Training dataset – Batch Evolution
3. Test dataset – Batch Summary
4. Test dataset – Batch Evolution

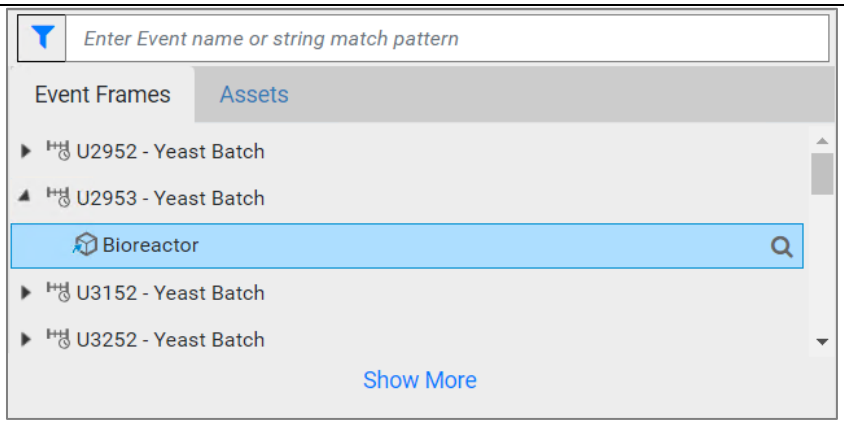
Step 1: Training dataset – Batch Summary

Select Data

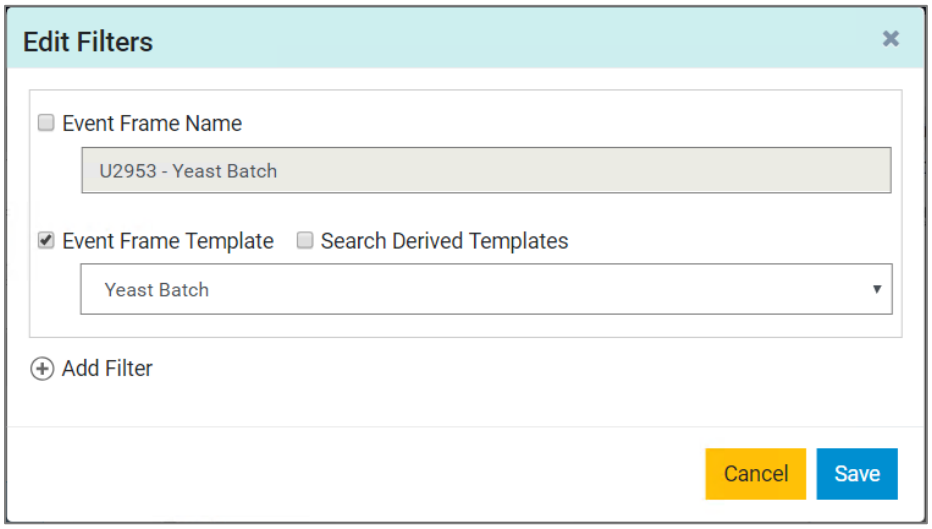
1. Open Google Chrome and click on the "PI Integrator for BA" bookmark (https://pisrv01:444/).
2. Click on "Create Event View".
3. Name the view "Modelling dataset – Batch Summary".
4. Click on "Create a New Shape".
5. On the Event View page, select the Database "Chemicals".
6. Drag and drop on the of Yeast Batch Events into the event shape.
7. In the attributes pane at the bottom-left of the screen, press the sorting button and select "Group by Category".
8. Drag and Drop the "Batch Info" and "Lab Data" categories into the event shape.

Note: We will not bring in the "Process Data" attributes (e.g. Max Ethanol) in the event shape. The integrator does not support the "By Time Range" setting of Event Frame attributes. To get the Max Ethanol concentration, we will bring in the Bioreactor's "Ethanol" attribute, and specify the aggregation needed in the next step.

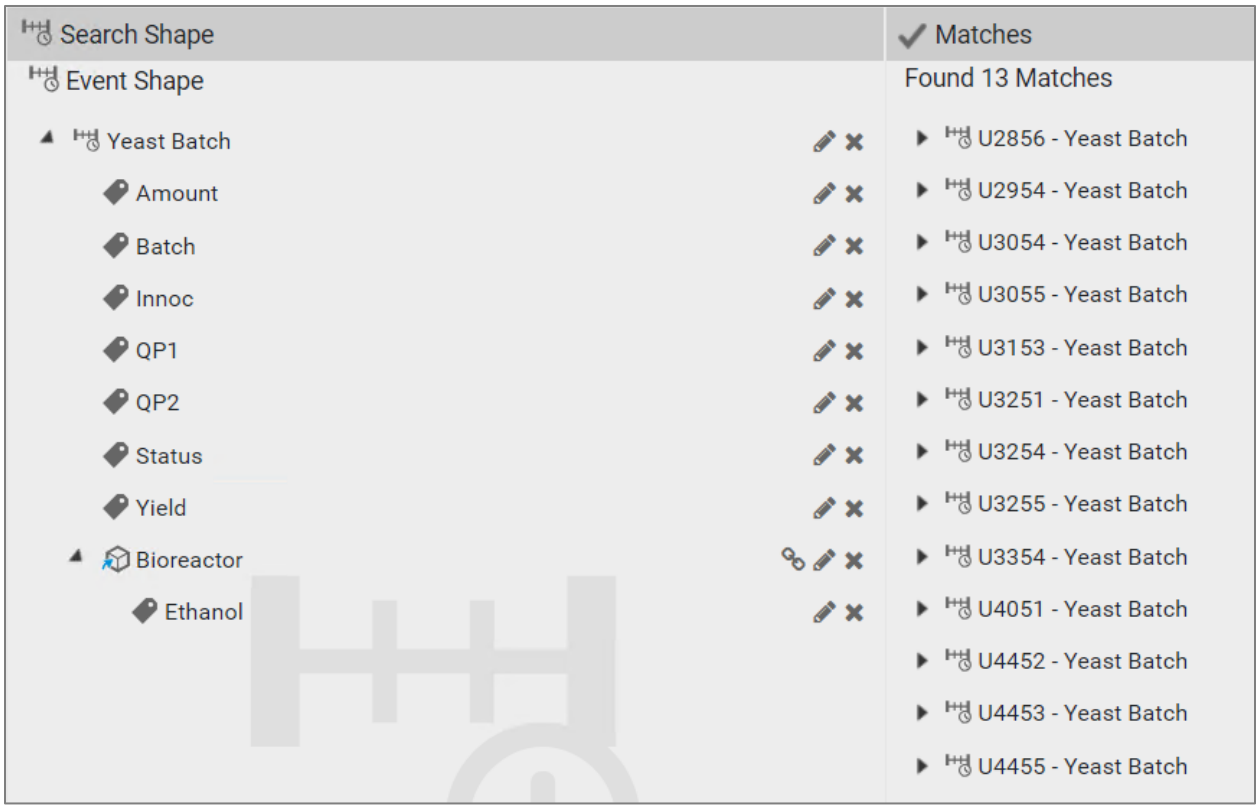
9. In the Event Frames pane, expand the select event. You should see the "Bioreactor" element. Click on it:



10. In the attributes pane at the bottom-left of the screen, drag and drop "Ethanol" into the event shape
11. Click on the edit button next to the batch you dragged into the event shape.
12. Uncheck "Event Frame Name" and check "Event Frame Template".



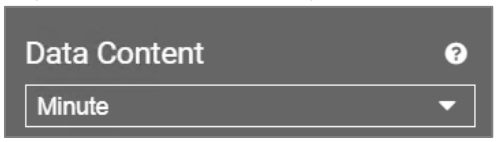
13. Your final result should look like the image below. If it does, click "Next".



Modify View

1. On the "Modify View" page, change the start time to "1-dec-2018" and the end time to "15-dec-2018"
2. Remove the "Timestamp" and "Bioreactor" columns
    - Select the column
    - In the right-hand pane, click the "Remove Column" button

      .
3. Change the "Event Frame Duration" Column from hours to minutes.
    - Select the column
    - In the right-hand pane, under "Data Content", select "Minute".

    - Click "Apply Changes"

4. For the "Ethanol" attributes, we must specify that we want to aggregate the data for the duration of the Event Frame.
    - Select the column
    - In the right-hand pane, under "Data Content", select "Maximum"
    - Click "Apply changes"

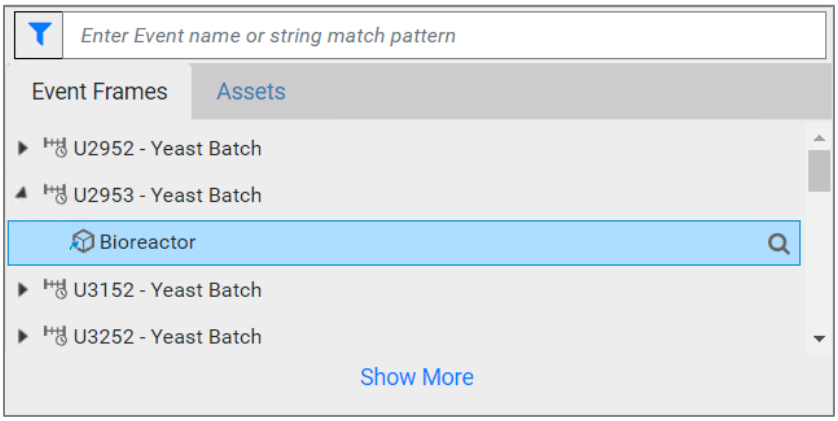
The data should look like the table below. If it does, click "Next".

1. On the “Publish” page, under “Target Configuration” select the target named “CSV”
2. Leave the Run Mode as “Run Once”
3. Click “Publish”

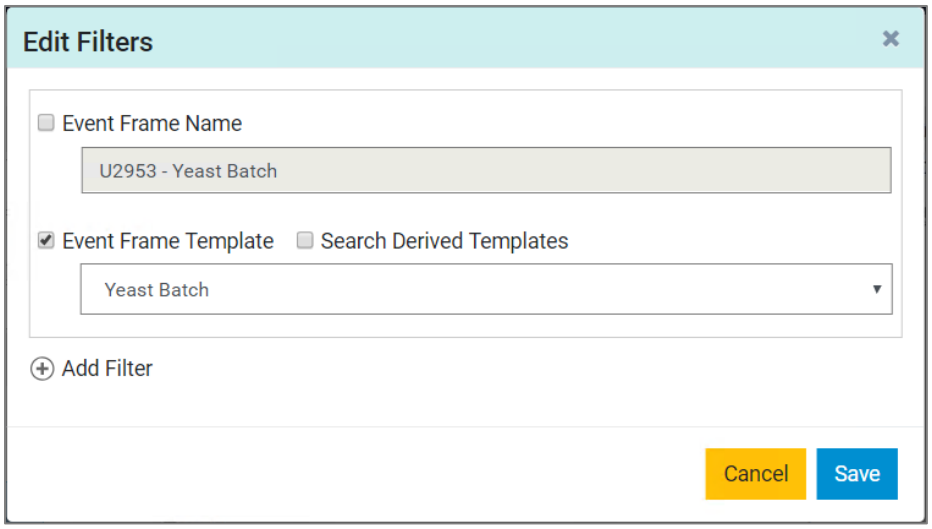
Step 2: Training dataset – Batch Evolution

Select Data

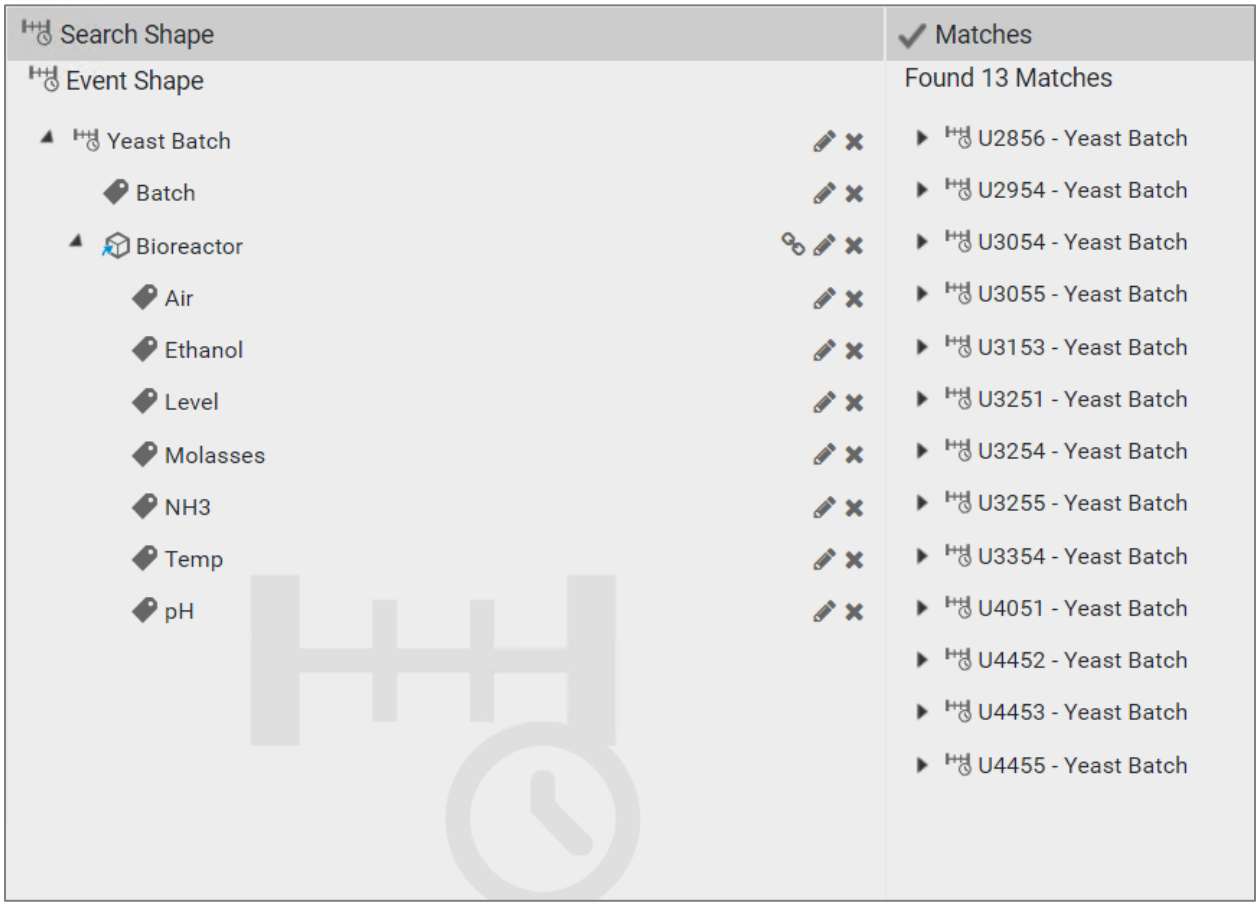
1. On the PI Integrator homepage, click on “Create Event View”.
2. Name the view “Modelling dataset – Batch Summary”.
3. Click on “Create a New Shape”.
4. On the Event View page, select the Database “Chemicals”.
5. Select on the of Yeast Batch Events.
6. From the attributes pane, drag and drop the “Batch” attribute into the Event Shape.
14. In the Event Frames pane, expand the select event. You should see the “Bioreactor” element. Click on it:



7. In the attributes pane at the bottom-left of the screen, press the sorting button and select “Group by Category”.
8. Drag and Drop the “Process Data” category into the Event Shape
9. Click on the edit button next to the batch you dragged into the event shape.
10. Uncheck “Event Frame Name” and check “Event Frame Template”.

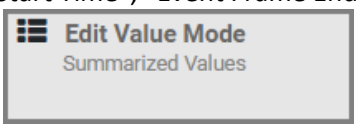


11. Your final result should look like the image below. If it does, click “Next”.

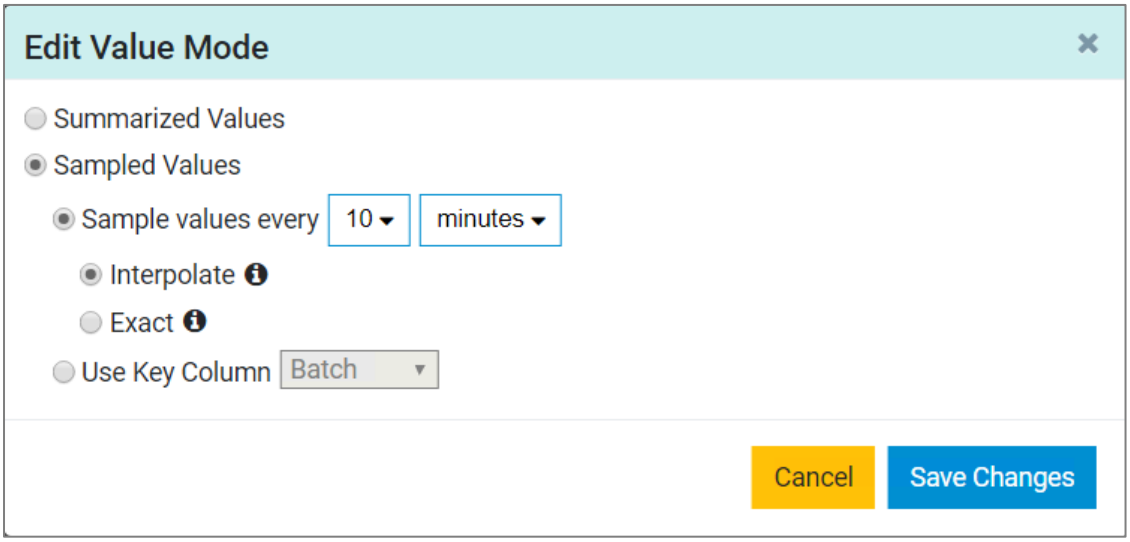


Modify View

1. On the “Modify View” page, change the start time to “1-dec-2018” and the end time to “15-dec-2018”
2. Remove the “Event Frame Start Time”, “Event Frame End Time” and “Duration” column
3. Click on “Edit Value Mode”



4. Select “Sample Values” and set “Sample values every 10 minutes”, with interpolation.



5. Click next.

Publish

1. On the “Publish” page, under “Target Configuration” select the target named “CSV”
2. Leave the Run Mode as “Run Once”
3. Click “Publish”

Step 3: Test dataset – Batch Summary

1. On the PI Integrator homepage, select the view you just created: “Modelling dataset – Batch Summary”.
2. Click “Modify View”

3. Select “Edit a copy of this View” and name it “Test dataset – Batch Summary”
4. On the “Select Data” page, click “Next”.
6. On the “Modify View” page, change the start time to “28-feb-2019” and the end time to “8-mar-2019”. Click “Next”
7. On the “Publish” page, click “Publish

Step 4: Test dataset – Batch Evolution

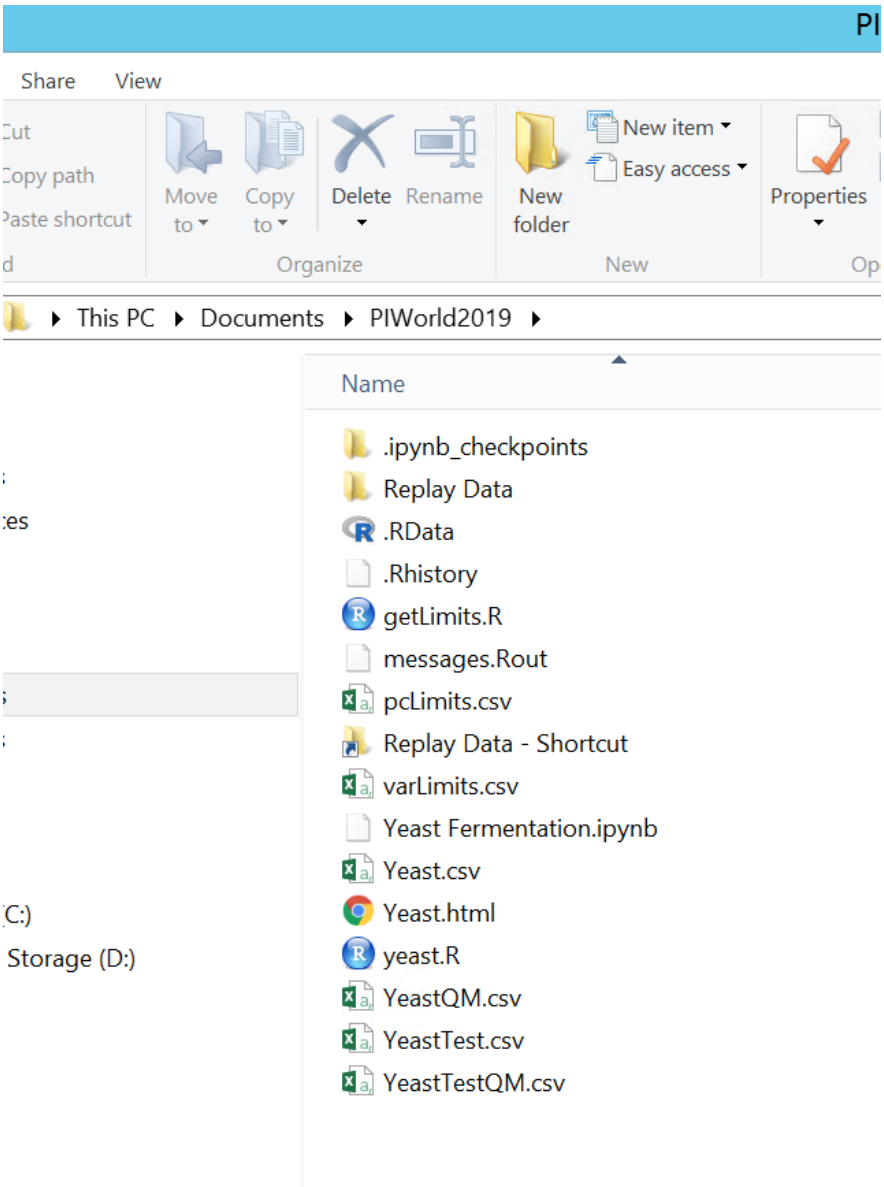
5. On the PI Integrator homepage, select the view you just created: “Modelling dataset – Batch Evolution”.
6. Click “Modify View”

7. Select “Edit a copy of this View” and name it “Test dataset – Batch Evolution”
8. On the “Select Data” page, click “Next”.
8. On the “Modify View” page, change the start time to “28-feb-2019” and the end time to “8-mar-2019”. Click “Next”
9. On the “Publish” page, click “Publish

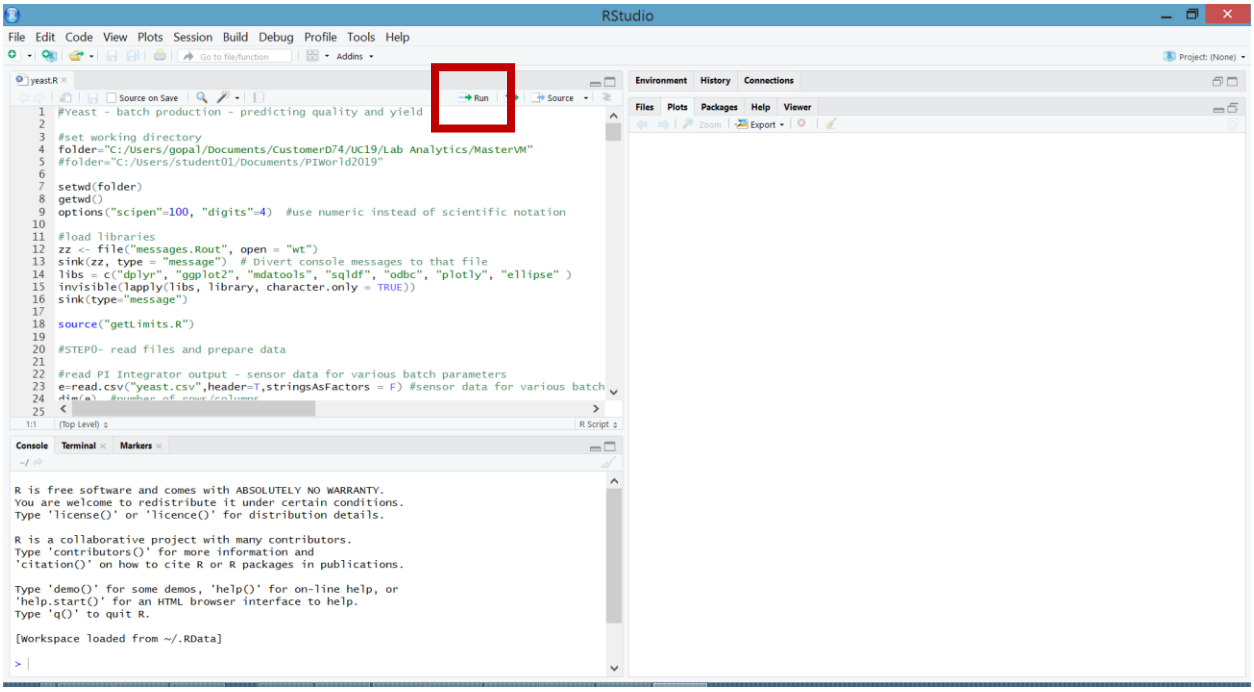
Part II – Model Development Using R

Exercise 1 – Monitoring the Batch Evolution

From the PIWorld2019 folder, select yeast.R and double-click to open it in R Studio (please be patient, R Studio takes a few seconds to open).

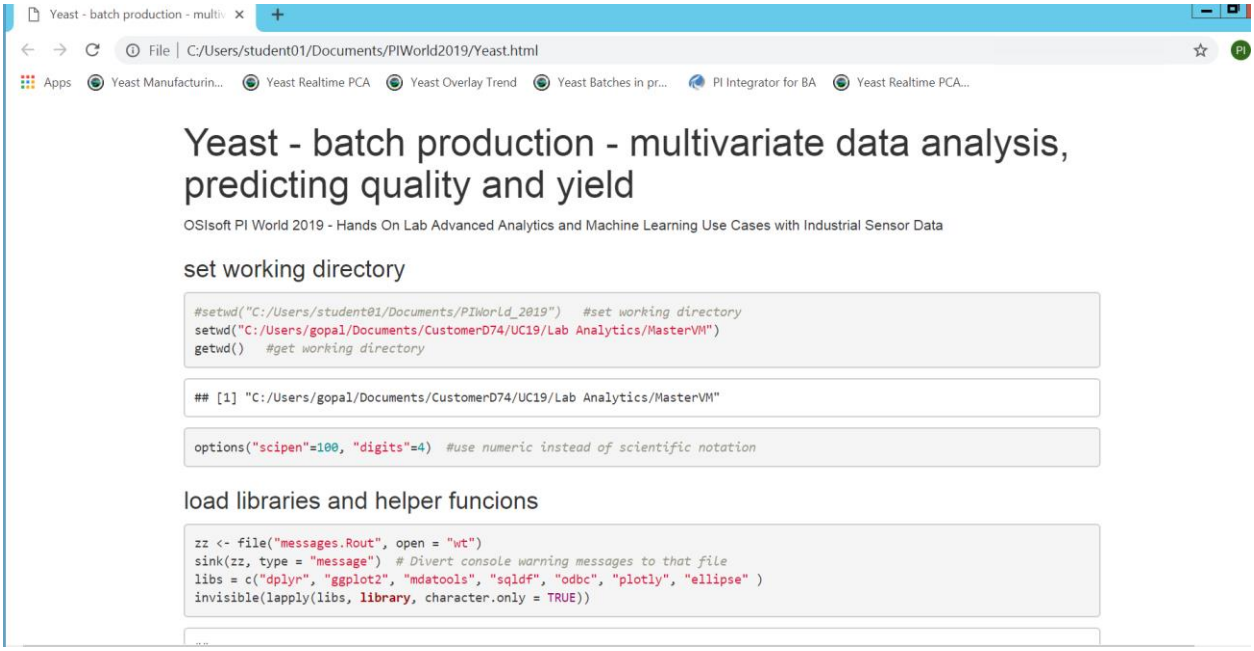


The screen below shows the R Studio user interface.



The output from the script as you run it line by line using **Run** is shown in Yeast.html (see the lab folder for the latest revision).

Only the first page from Yeast.html is shown below.



Please refer to the electronic copy of Yeast.html in the Lab VM.

Part II - Model deployment and real-time scoring

Step 1: AF Analytics

The relevant step from the R script is shown below (Step 1-5 in the R script output – html document)

Exercise 1 BATCH EVOLUTION ### STEP 1-5 deploy in AF Analytics

```
pc1eq=""  #get pc1 equation
for (j in 1:7)
{pc1eq=cat(sep="",pc1eq,"+(","'",names(egood.pca$center[j]),"'","-(",egood.pca$center[j],")",")/",egood.pca$scale[j],"*",egood.pca$loadings[j,1])}
```

```
## +('Ethanol'-(0.6546))/0.4914*0.3986+('Temp'-(31.16))/1.226*-0.4698+('Molasses'-(2136))/957.1*-0.06959+('NH3'-(107.5))/63.6*0.336+('Air'-(5872))/1444*-0.3142+('Level'-(48.63))/5.067*-0.5222+('pH'-(5.205))/0.3461*-0.3622
```

```
pc2eq=""  #get pc2 equation
for (j in 1:7)
{pc2eq=cat(sep="",pc2eq,"+(","'",names(egood.pca$center[j]),"'","-(",egood.pca$center[j],")",")/",egood.pca$scale[j],"*",egood.pca$loadings[j,2])}
```

```
## +('Ethanol'-(0.6546))/0.4914*-0.1048+('Temp'-(31.16))/1.226*0.1278+('Molasses'-(2136))/957.1*-0.6414+('NH3'-(107.5))/63.6*-0.5026+('Air'-(5872))/1444*-0.5366+('Level'-(48.63))/5.067*-0.1381+('pH'-(5.205))/0.3461*0.04061
```

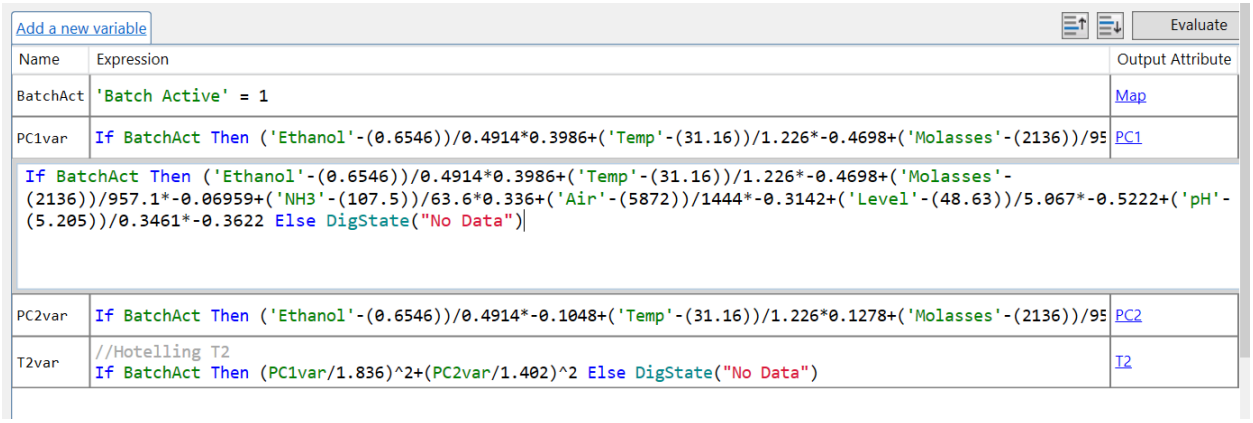
```
#T2 equation
T2eq=""
T2eq=cat(sep="",'(pc1/',sd(egood.pca$calres$scores[,1]),')^2+(pc2/', sd(egood.pca$calres$scores[,2]) ,')^2')
```

```
## (pc1/1.836)^2+(pc2/1.402)^2
```

```
#Q equation
Qeq=""
for (j in 1:7)
{Qeq=cat(sep="",Qeq,"+(('",names(egood.pca$center[j]),"'","-(",egood.pca$center[j],")",")/",egood.pca$scale[j],"-(","PC1Var*",egood.pca$loadings[j,1],"+PC2Var*",egood.pca$loadings[j,2],"))^2")}
```

```
## +(('Ethanol'-(0.6546))/0.4914-(PC1Var*0.3986+PC2Var*-0.1048))^2+(('Temp'-(31.16))/1.226-(PC1Var*-0.4698+PC2Var*0.1278))^2+(('Molasses'-(2136))/957.1-(PC1Var*-0.06959+PC2Var*-0.6414))^2+(('NH3'-(107.5))/63.6-(PC1Var*0.336+PC2Var*-0.5026))^2+(('Air'-(5872))/1444-(PC1Var*-0.3142+PC2Var*-0.5366))^2+(('Level'-(48.63))/5.067-(PC1Var*-0.5222+PC2Var*-0.1381))^2+(('pH'-(5.205))/0.3461-(PC1Var*-0.3622+PC2Var*0.04061))^2
```

Step 2: (Optional) Copy/paste the equations from R to AF Analytics

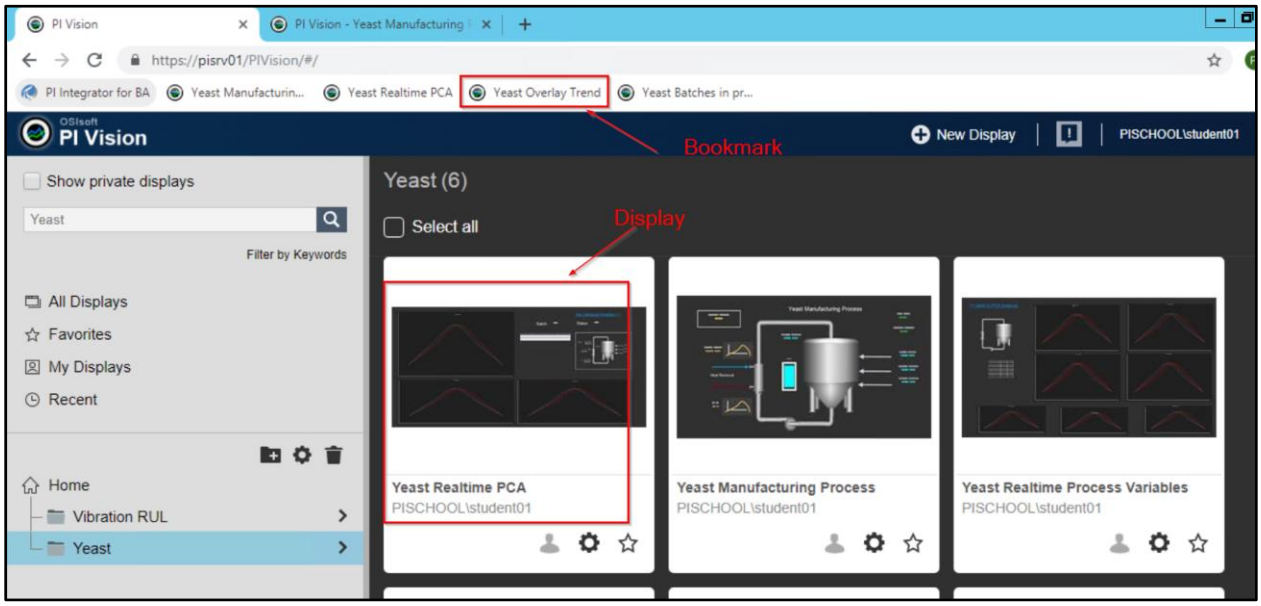


Step 3: Real-time scoring

To mimic the user experience during an actual batch production run, replay the data from one of the batches (U3054) as shown below.

You can configure a PI Notification to receive an email with a hyperlink to a PI Vision display to further diagnose the problem.

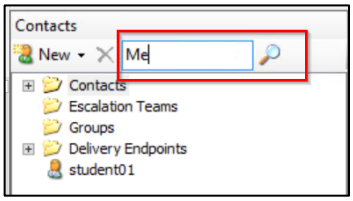
Alternately, you can open the ***Yeast Realtime PCA*** display in PI Vision – use the appropriate bookmark in Google Chrome.



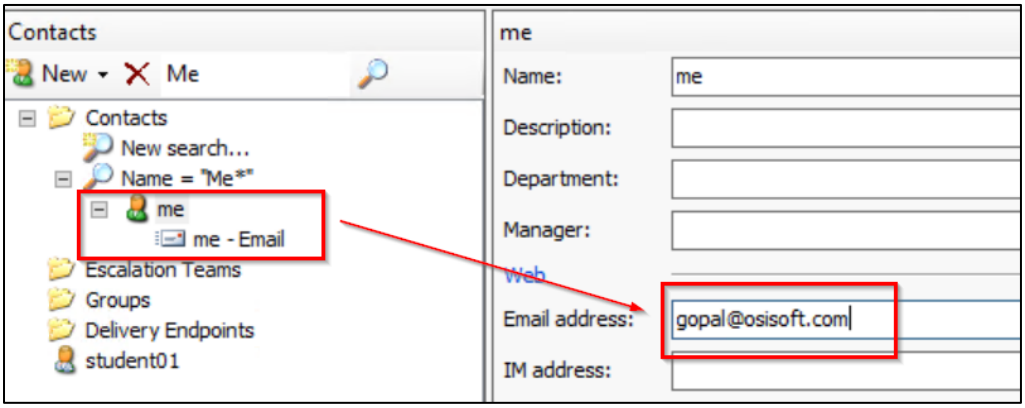
Configure PI Notifications to receive an email (optional)

1. In PI System Explorer, navigate to Contacts tab:

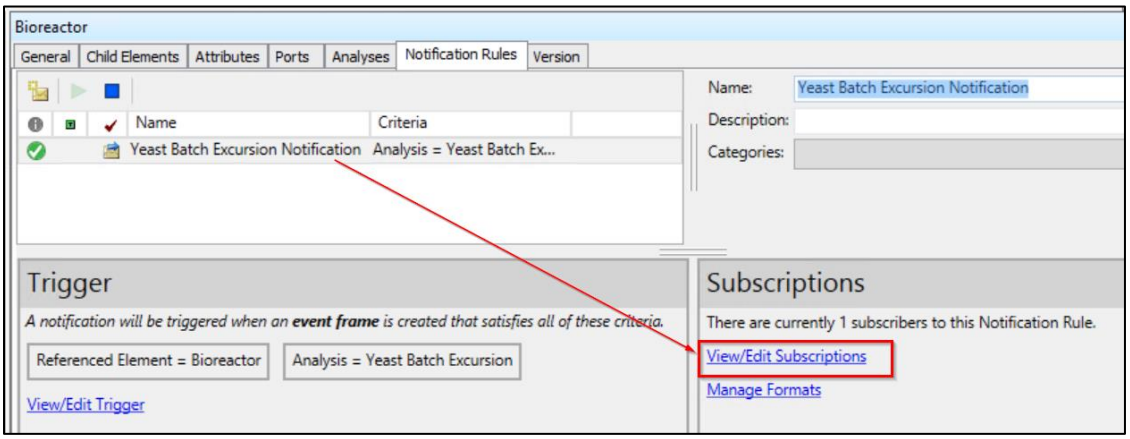
2. In the Search box, type ***Me*** and click on the icon to search



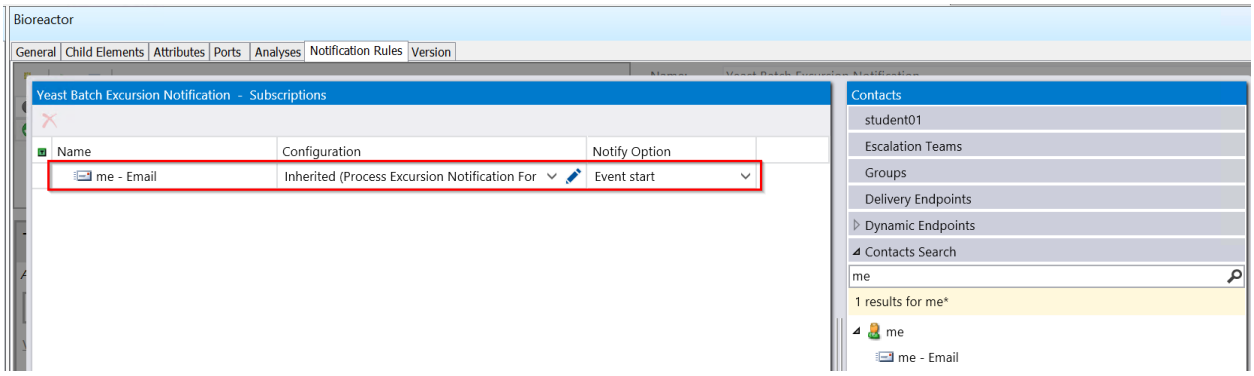
3. In the results, select the contact Me, and change the email address to your own:



4. You can verify that you are subscribed to the notification as follows:
    a. In the ***Elements*** tab, select the ***Bioreactor*** element and go under ***Notification Rules***
    b. Select **Yeast Batch Excursion Notification** and click the View/Edit Subscriptions link

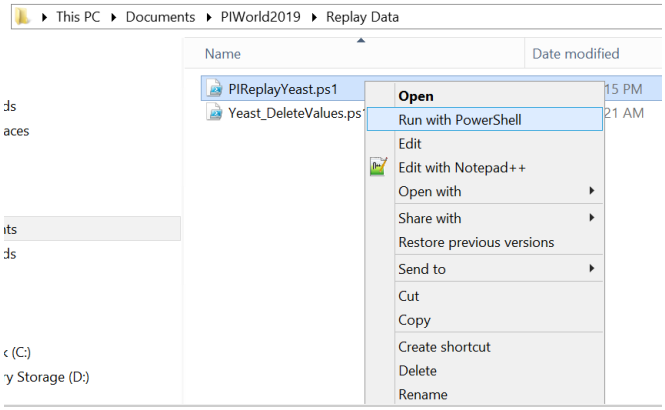


    c. Verify that the contact you just modified (Me) is in the subscribers list:



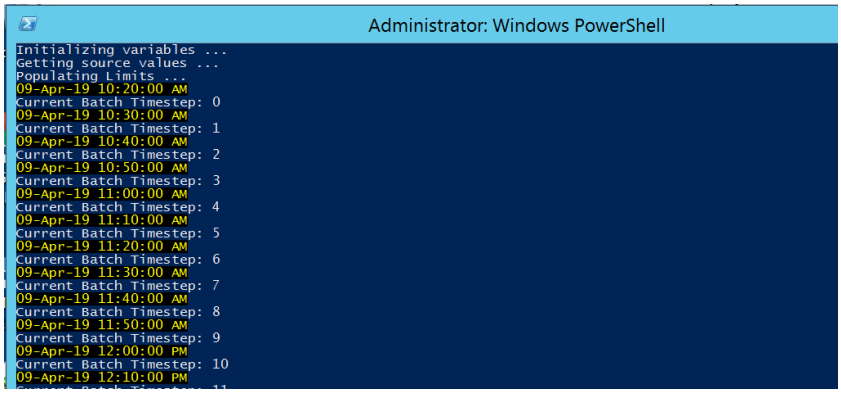
Replay the data for Batch U3054

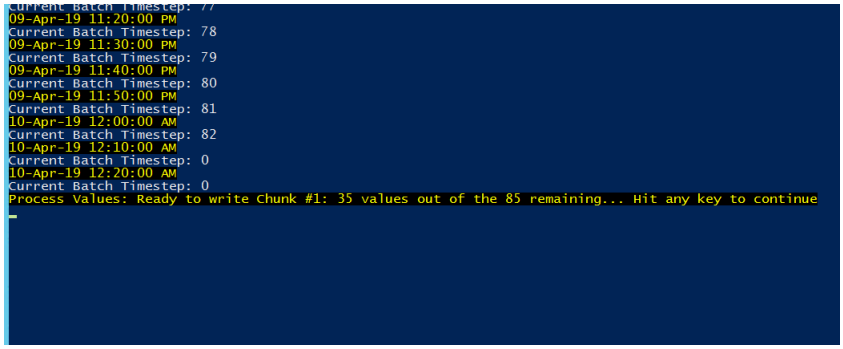
1. Right Click **PIReplayYeast.ps1** in <Replay Data> folder to bring up the context menu, and select “Run with PowerShell”



If you receive a Warning message regarding Execution Policy etc., say “Y” to continue.

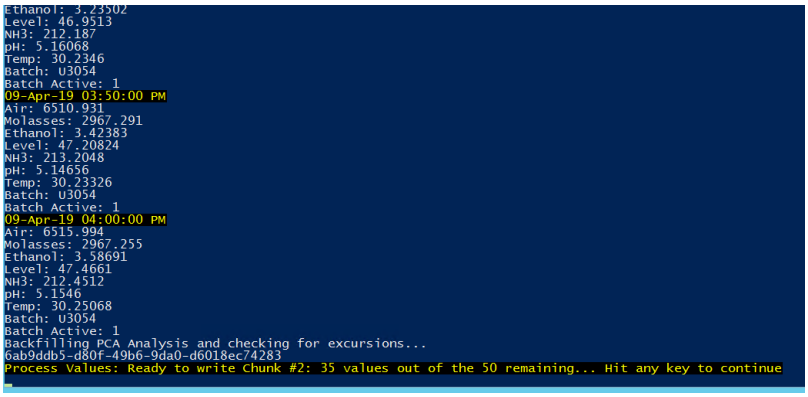
2. The script will initialize variables and prepopulate the process limits.





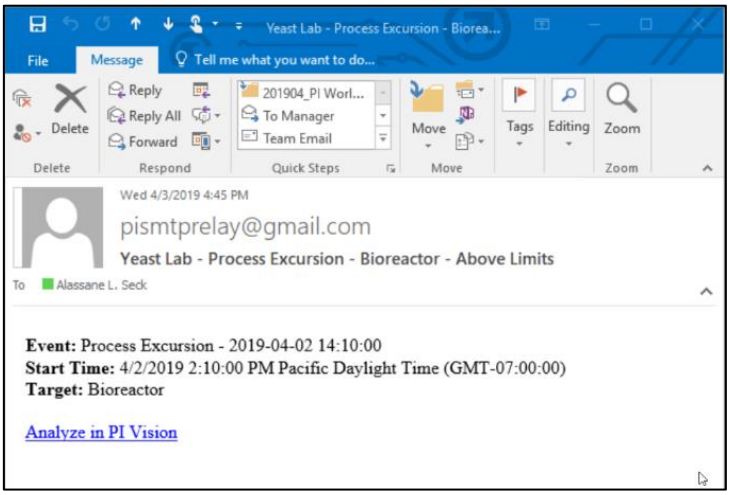
Once it reaches the step shown, it is ready to fill the real-time process values. Hit any key to fill the first chunk of values.

After the first several values of each process variable are written, the script will pause.



If you configured PI Notifications to send you an email, you should receive one at this point.

Note that the PI Notification alerts have been configured against T2 and Q limits and not against the individual process variables.

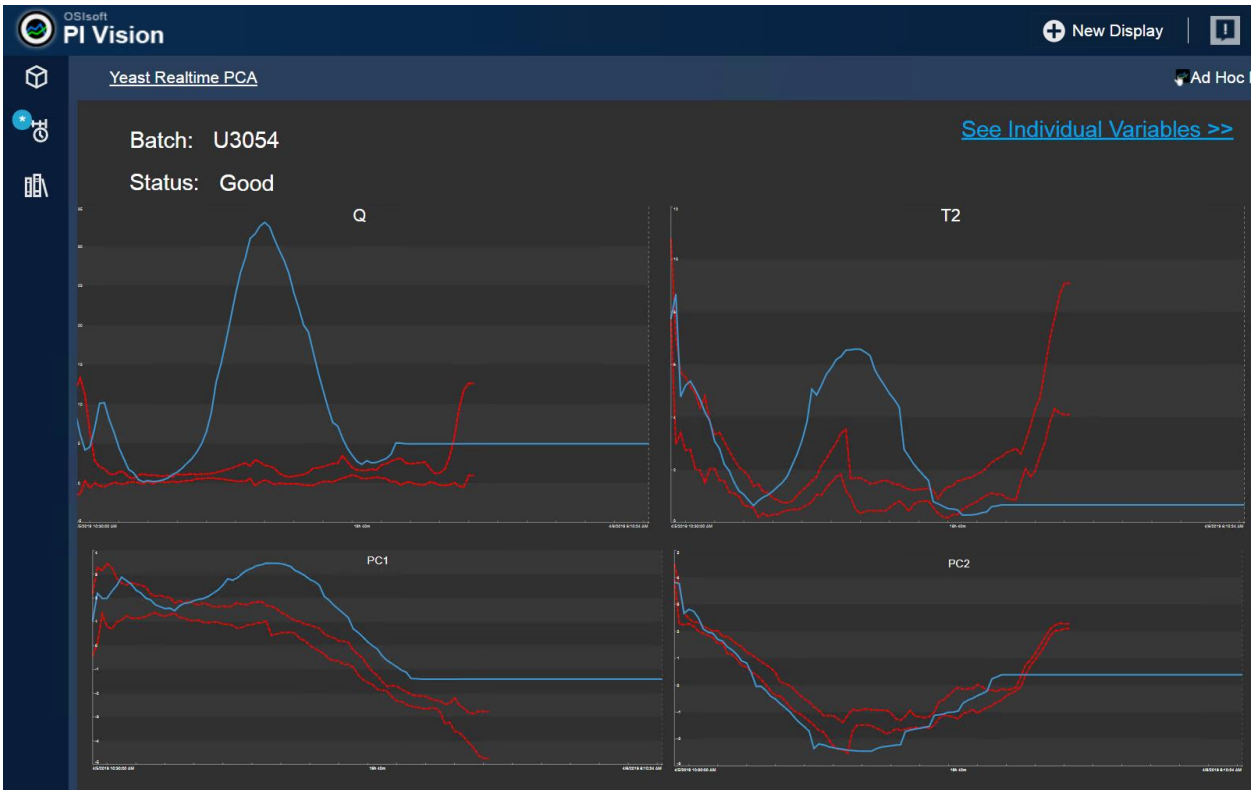


You can also directly navigate to the ***Yeast Realtime PCA*** display in PI Vision and observe the evolution of the PCA variables.

The latter should clearly show that your process went out of control.

You can navigate to the display showing individual variables to assess which variable went out of limits exactly. For this batch run, it is clear the ethanol exceeded acceptable limits.

3. Play the second chunk of data where the ethanol back under control limits.



Part III – Model Development Using R – Full Batch Assessment

Exercise 2 Full Batch Assessment

In this exercise, we prepare the data with time in columns instead of rows for full batch assessment.

See Yeast.html for R script and output.

Part IV – Model Development Using R – Predict Quality

Exercise 3 Predict Quality

In this exercise, we use the data format from Exercise 2 but with PLS to predict quality.

See Yeast.html for R script and output.

Reference Materials

https://cran.r-project.org/

https://shiny.rstudio.com/

http://stackoverflow.com/questions/22309236/options-for-deploying-r-models-in-production

Yeast dataset: https://landing.umetrics.com/downloads-other-downloads

Have an idea how to improve our products?
OSIsoft wants to hear from you!
https://feedback.osisoft.com/