Bioprocess real-time prediction

Demonstrating Waters LS-MC and JMP analytics for Smarter Bioprocess

Russel Wolfinger, Director of Scientific Discovery and Genomics, Distinguished Research Fellow
Hyewon Jung, Sr. Associate Systems Engineer

Introduction of Waters Add-in

In upstream bioprocessing, the **LC-MS (Liquid Chromatography–Mass Spectrometry)** system is used to separate and quantify dozens to hundreds of chemical compounds contained in culture samples — such as metabolites, proteins, and media component

Waters **BioAccord LC-MS** can be directly connected to automated bioreactors such as **Amber 15** or **Amber 250**.
This allows process engineers to collect and analyze LC-MS data directly from the bioreactor — without waiting for an external analytical lab.
As a result, engineers can monitor metabolites and product attributes in **near real time** during the culture run.

To fully leverage these data, JMP and Waters jointly developed the Bioprocess Monitor Add-In,
which enables seamless data integration, visualization, and advanced modeling directly within JMP.

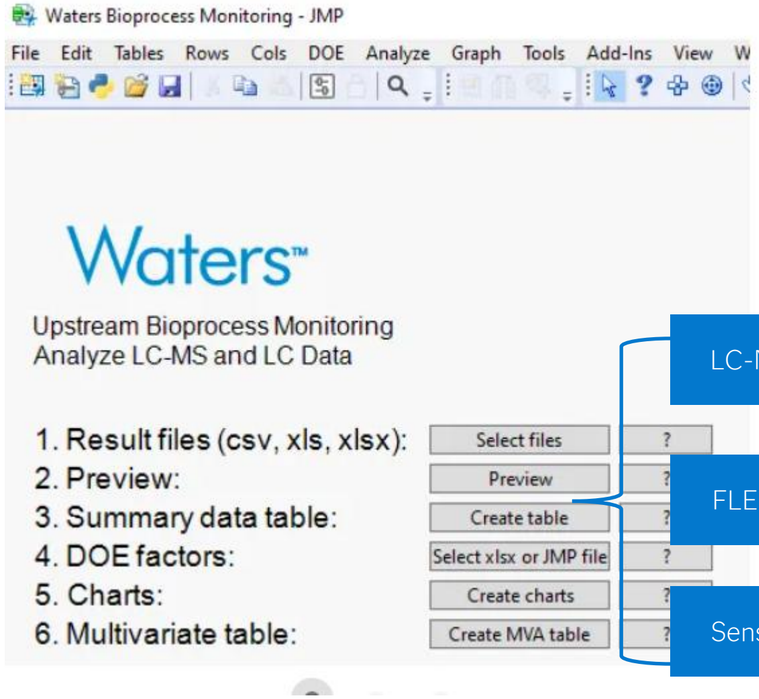
Data structure of BioAccord

1. **대용량 · 다차원 데이터**
한 번의 LC-MS 실험에서도 **수천~수만 행**(row) 의 데이터가 생성됨
예시:
8개 바이오리액터 × 14일 배양 × 하루 100개 이상 분석 항목
→ **약 12,000행 이상의 데이터가 한 번의 실험에서 생성**

2. **열**(Column) **구성**
메타정보(Metadata)
Bioreactor ID, Day, Sample ID, 측정 장비, 측정 시점 등
Analyte **값**
각 화합물(Analyte)의 정량 결과
추가 속성
공정 조건, 측정 환경 등의 보조 데이터

3. **데이터 형태**
Tall Format(**세로형 구조**)
시간(시계열), 화학(Analyte), 공정 변수 등
다차원 데이터를 “세로로” 쌓은 형태
→ Waters LC-MS **데이터는 다차원 시계열·화학·공정 데이터**

Introduction of Waters Add-in



이 데이터는 매우 풍부하지만,
그만큼 **정리·통합·시각화가 어렵다**는 문제가 있습니다.
게다가 같은 실험에서도 LC-MS, FLEX2, 센서 등 여러 장비의 결과를 합쳐야 하기 때문에 어려움이 따릅니다.

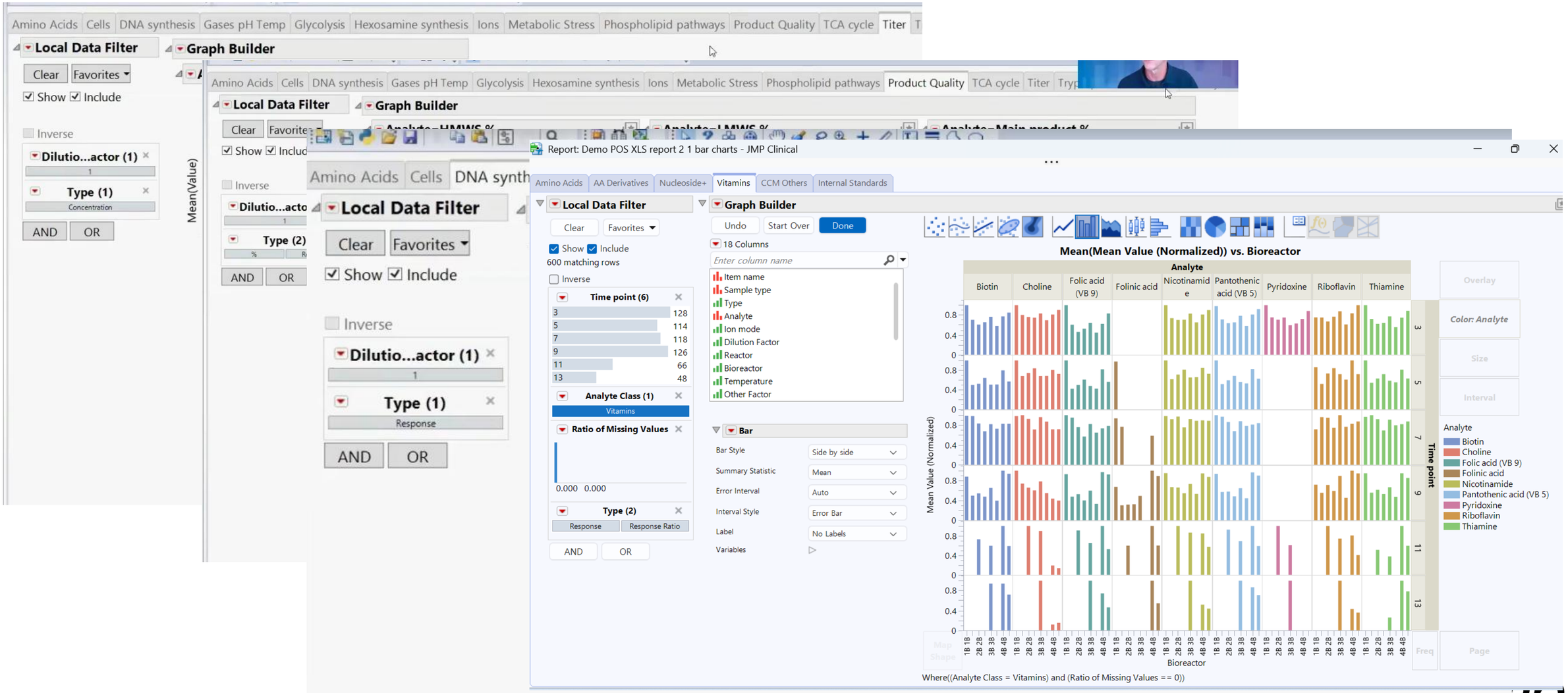
이 장비들로부터 모두 얻어지는 데이터의 구조는 다음과 같습니다.

- **제품 품질 지표**: Productivity (Titer), Aggregates and Fragments(%Main Product, %HMWS, %LMWS), Glycoforms (%Afucosylation), Cell Performance (Viability, density)
- **대사체 데이터**: 아미노산, 비타민, TCA 사이클 중간체 , Glycolysis Fulux (sugar utilization), Oxidative Stress(Cellular health)
- **세포 대사 정보**: 영양소 소비 패턴, 부산물 생성량 등
- **공정 입력 변수**: Process Conditions (Media, pH, DO, Temp, Reactor type, Agitation…), Clone selection (cell line characteristics), Feeding strategy (Nutrient timing &composition)— Gas flow rates (AIR/O2 sparge rates)

-> 그래서 Waters와 JMP가 공동으로 개발한 Bioprocess Monitor Add-In이 등장했습니다.
이 Add-In은 이러한 Waters LC-MS 데이터를 자동으로 불러오고, 메타데이터를 추가하며,
시각화와 다변량 분석이 가능한 형태로 정리해줍니다.

Waters Add-in Output

1) Visualization (막대그래프 or 선그래프)



Waters Add-in and Further Analysis

2) MVA 테이블 생성 -> Starting point for further analysis

Understanding Bioprocess Complexity: Advanced Analytical Profiling with JMP
Discovery Summit Austin 2025

1. Exploratory Visualization and Response Screening

— Greg Flexman

Visual exploration of variable relationships using **Graph Builder** and identification of key influencing factors through **Response Screening**.

2. Functional Data Analysis (FDA)

— Ben Barroso-Ingham

Investigation of dynamic relationships between **sensor data** and **DOE factors**, uncovering time-dependent effects in bioprocess profiles.

3. Final Titer Prediction using Mixed Model & XGBoost

— Russ Wolfinger

Application of **XGBoost modeling** to predict **final yield (titer)** based on multivariate process data and sensor-derived features.

Batch Data Analysis in Bioprocessing

From Monitoring to Real time prediction

Why Do It?

- Even under the same process, each batch varies due to sensor drift, raw material differences, or environmental fluctuations.
- Traditional QC detects issues only after production — a reactive, delayed approach.
- Batch data analysis enables real-time monitoring and predictive control, connecting in-process data directly to decision-making.

In Short:

These analysis tools establish the foundation for implementing a Process Analytical Technology (PAT) protocol — ensuring continuous, data-driven process understanding and control.

Batch Data Analysis in Bioprocessing

Statistical & Machine Learning Methods Used

Multivariate Control & Outlier Detection

Methods: PCA, FDE Control chart, Multivariate Control Chart

- *Detect abnormal batches or sensor drifts in real time*
- *Visualize batch-to-batch variation and process stability*
- *Enable early warning and deviation monitoring*

Predictive Modeling

Methods: PLS, FDE, GenReg, XGBoost, Neural Networks, PyTorch

- Predict final titer from multivariate process data
- Identify key influencing factors and latent patterns
- Support real-time, data-driven process control

Data Structure

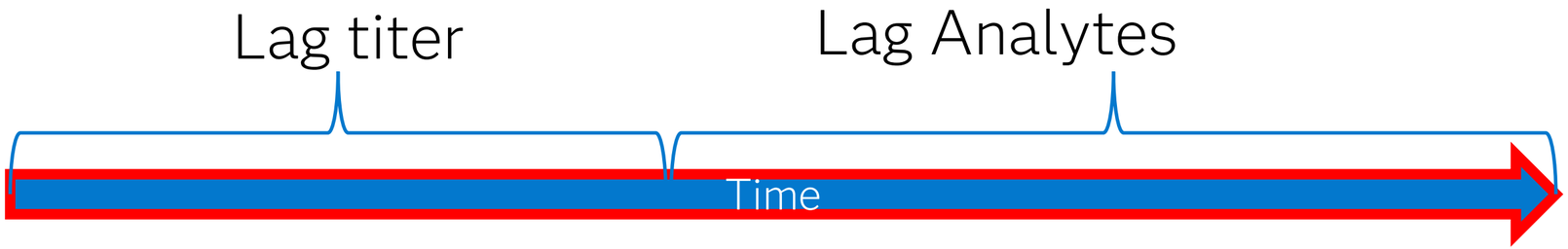
DOE Factor

Products, Vitamin, Urea cycles, synthesis, metabolism, TCA Cycle…

Time

Bioreactor	Time point	Concatenate[Bioreactor,Time point]	Factors			Product							Vitamins				
			Air Overlay Rate	Air Sparge Rate	Buffer Concentration	%	LMWS %	Main product %	Main product %*Titer	OLD Titer	Titer Increment	Maximum[Titer]	Folic acid (VB 9)	Nicotinamide	Nicotinic acid	Pantothenic acid (VB 5)	Thiamine
V1	0	V1_D0	0	0.5	10	39.8	70.6	100	1.33	2.03	1.8		3721	135k	2776	20.4k	56.9
V2	1	V1_D1	50	5	22.6												
V3	2	V1_D2				0	0	29.4	0.011	0.16	1.06		1839	74.8k	1185	10.2k	33.8
V4	3	V1_D3															
V5	4	V1_D4															
3 others	10 others	115 others															
V1	0	V1_D0	0	5	22.6	0.0	70.6	29.4	0.0114709964	0.179	.		.	.	.	.	.
V1	1	V1_D1	0	5	22.6	.	.	.	.	.	1.1866348554						
V1	2	V1_D2	0	5	22.6	0.0	67.5	32.5	0.0178841442	0.209	1.1866348554		3167.5	125148.3	2198.5	20383.4	51376.3
V1	3	V1_D3	0	5	22.6	.	.	.	.	.	1.3250804445		2938.0	118771.8	2061.1	18836.7	46024.8
V1	4	V1_D4	0	5	22.6	0.0	50.8	49.2	0.0474606665	0.277	1.3250804445		2902.7	115790.2	1984.1	18798.7	45853.0
V1	5	V1_D5	0	5	22.6	.	.	.	.	.	1.6116015664		2989.7	118403.2	2045.0	19288.5	47649.5
V1	6	V1_D6	0	5	22.6	24.2	19.4	56.4	0.1412395042	0.278	1.6116015664		3396.6	121087.5	2075.1	19423.5	49119.7
V1	7	V1_D7	0	5	22.6	22.6	18.5	58.8	0.1996378636	0.364	1.3541369431		3403.8	117670.4	2136.1	18568.7	49146.1
V1	8	V1_D8	0	5	22.6	20.8	15.5	63.7	0.3516375095	0.572	1.6274802561		2999.9	112205.3	1914.9	16924.0	46910.8
V1	9	V1_D9	0	5	22.6	18.9	16.5	64.6	0.5161945844	0.826	1.4465197713		2855.3	105146.4	2062.6	15612.9	39876.3
V1	10	V1_D10	0	5	22.6	15.5	20.4	64.1	0.6027889236	0.942	1.1770380506		2707.1	100078.3	1865.5	12905.4	38345.6
V1	11	V1_D11	0	5	22.6	15.2	19.3	65.6	0.7804056957	1.192	1.2661335906		2726.6	98406.3	1752.9	12916.2	39735.0
V1	12	V1_D12	0	5	22.6	16.0	16.2	67.9	0.9129323107	1.349	1.1302411914		2150.4	86452.8	1440.7	12874.0	40484.1
V1	13	V1_D13	0	5	22.6	16.1	16.0	67.8	0.973346736	1.441	1.0664461599		.	.	.	.	.
V1	14	V1_D14	0	5	22.6	16.6	15.4	68.1	1.0318380775	1.515	1.0563436121		2159.7	74765.2	1344.3	15527.6	41526.5
			0	5	22.6	0.0	0.0	66.7	0.0215027276	0.168	.		.	.	.	.	.
													.	.	.	.	.

Batch Level Data Structure (Wide Structure)



			Lag Titer											Lag Analytes											
Bioreactor	Titer 14	Predicted Titer 14	Titer 0	Titer 2	Titer 4	Titer 6	Titer 7	Titer 8	Titer 9	Titer 10	Titer 11	Titer 12	Titer 13	HMW S % 0	HMW S % 1	HMW S % 2	HMW S % 3	HMW S % 4	HMW S % 5	HMW S % 6	HMW S % 7	HMW S % 8	HMW S % 9	HMW S % 10	HMW S % 11	HMW S % 12	HMW S % 13
V1	2.01	1.9	0.039	0.055	0.096	0.27	0.37	0.63	0.92	1.16	1.46	1.7	1.86					27.6		39.8	36.4	29.4	25.4	22	20.5	19.6	19.1
V2																											
V3																											
V4																											
V5																											
3 others	1.36	1.4	0.026	0.04	0.07	0.18	0.25	0.45	0.66	0.8	1.02	1.19	1.28	0		0		0		20.7	20.8	19.3	16.4	14.6	13.7	13.8	14.7
V1	1.516	1.5214020393	0.039	0.055	0.096	0.251	0.339	0.552	0.799	0.940	1.190	1.345	1.435	0.0	•	0.0	•	0.0	•	24.2	22.6	20.8	18.9	15.5	15.2	16.0	16.1
V2	1.361	1.3996829701	0.032	0.049	0.094	0.228	0.304	0.534	0.686	0.843	1.025	1.191	1.282	0.0	•	0.0	•	27.6	•	30.7	29.9	26.2	24.6	22.0	20.5	19.6	19.1
V3	1.523	1.5274105786	0.028	0.041	0.082	0.226	0.311	0.517	0.730	0.927	1.144	1.317	1.418	0.0	•	0.0	•	0.0	•	28.7	28.1	24.0	21.4	17.8	16.4	16.5	16.7
V4	1.595	1.5818277013	0.027	0.041	0.083	0.217	0.291	0.506	0.713	0.920	1.134	1.348	1.478	0.0	•	0.0	•	27.4	•	33.1	33.2	27.7	25.4	22.0	20.2	18.6	18.0
V5	1.474	1.5537233351	0.030	0.040	0.070	0.183	0.253	0.454	0.663	0.798	1.081	1.280	1.368	0.0	•	0.0	•	0.0	•	23.0	22.7	21.5	19.1	16.1	14.8	13.8	14.7
V6	1.807	1.8038539214	0.030	0.041	0.082	0.234	0.325	0.547	0.797	1.027	1.280	1.508	1.649	0.0	•	0.0	•	0.0	•	39.8	36.4	29.4	24.6	21.3	19.5	18.7	18.6
V7	1.620	1.6232113315	0.026	0.040	0.085	0.245	0.340	0.579	0.828	1.023	1.276	1.422	1.506	0.0	•	0.0	•	0.0	•	20.7	20.8	19.3	16.4	14.6	13.7	15.2	15.2
V8	2.008	1.900912436	0.030	0.045	0.091	0.270	0.374	0.627	0.919	1.164	1.464	1.698	1.860	0.0	•	0.0	•	0.0	•	32.6	30.5	24.9	20.4	17.2	17.5	17.5	17.5

Every batch becomes a single row in the data table

Demo In JMP

Closing Slide

Website or email