

Многомерный статистический контроль процессов (MSPC)

Докладчик

Алексей Померанцев

Со-автор

Оксана Родионова



*Институт химической
физики РАН им Семенова*



*Российское хемометрическое
общество*

Содержание

1. Введение
2. SPC
3. Хемометрика
4. MSPC
5. Batch MSPC
6. Выводы
7. Информация

Контроль процессов в реальном времени

Главная проблема, с которой сталкивается современная промышленность – это обеспечение постоянно высокого качества конечного продукта в типовом производственном процессе.

Принимая во внимание увеличивающуюся глобальную конкуренцию и быстро меняющиеся потребности рынка, эффективный контроль процессов в реальном времени становится насущной потребностью всех производящих компаний.

Цитата

Process Variability

If you are responsible for the management or improvement of a process then you need to understand process variability.

‘If I had to reduce my message for management to just a few words, I’d say it all had to do with reducing variability.’

Dr W. Edwards Demming

a uk analytical partnership forum



Многомерный статистический контроль процессов (MSPC)

MSPC – это математический анализ реальных исторических данных, характеризующих опыт работы

- **Цель:** Научиться у самих себя принимать оптимальные решения в различных ситуациях.
- **Средства:** Сбор истории работы процесса, а также анализ накопленных данных.
- **Результат:** Снижению затрат при стабилизации качества.

MSPC – это один из методов анализа процессов (PAT)

PAT & FDA

Process Analytical Technology (PAT) =
Технология (методы) анализа процессов

PAT =
Статистический контроль процессов (SPC) +
Хемометрика (Chemometrics)

**FDA = U.S. Department of Health and Human
Services Food and Drug Administration**

Guidance for Industry PAT — A Framework for Innovative
Pharmaceutical Development, Manufacturing, and Quality Assurance
Pharmaceutical CGMPs, September 2004

Статистический контроль процессов (SPC)

Цель

Контроль хода процесса

Концепция

Анализ накопленных исторических данных о процессе

Методы

Традиционные статистические методы

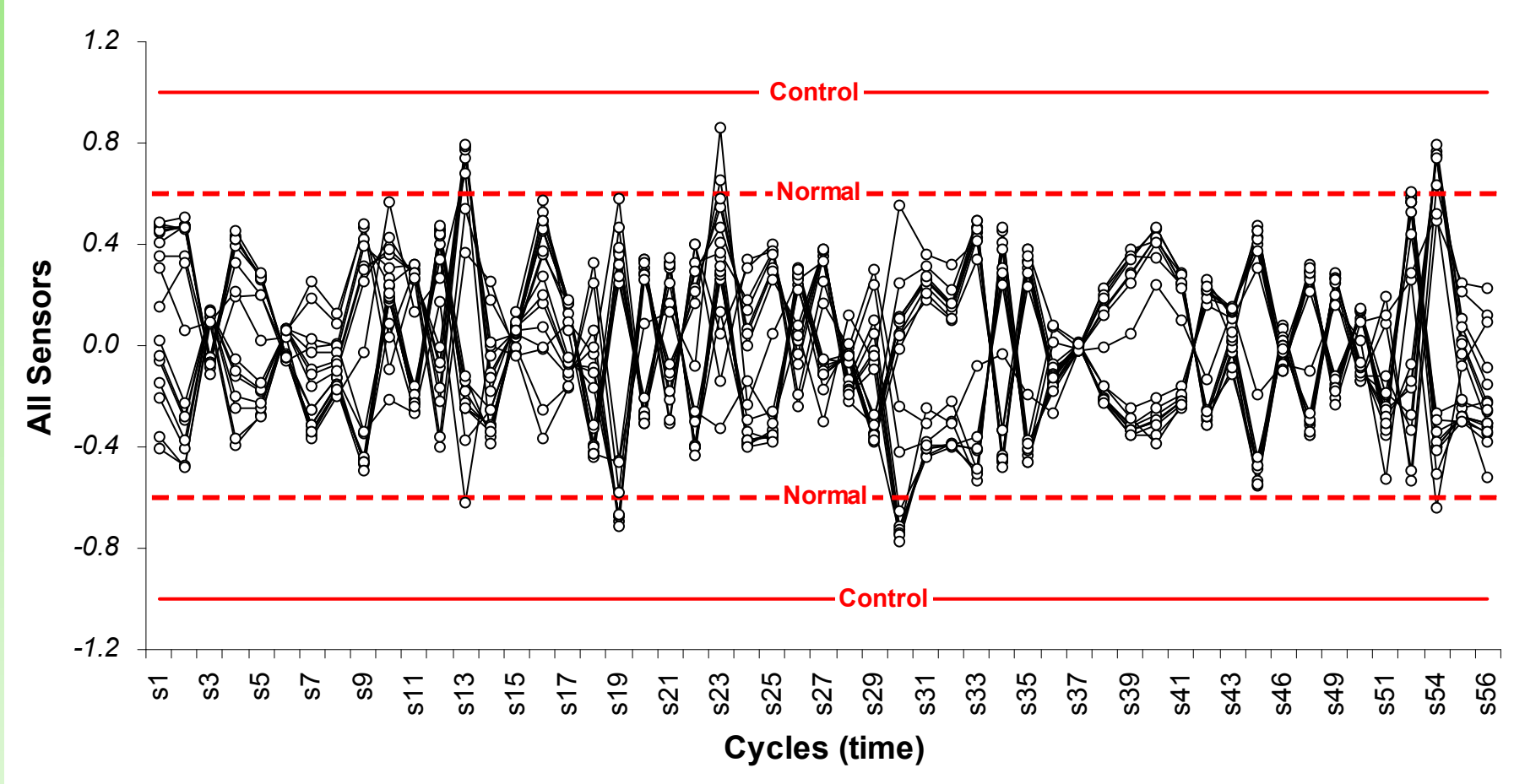
Реализация

Построение одномерных графиков – карт для всех датчиков

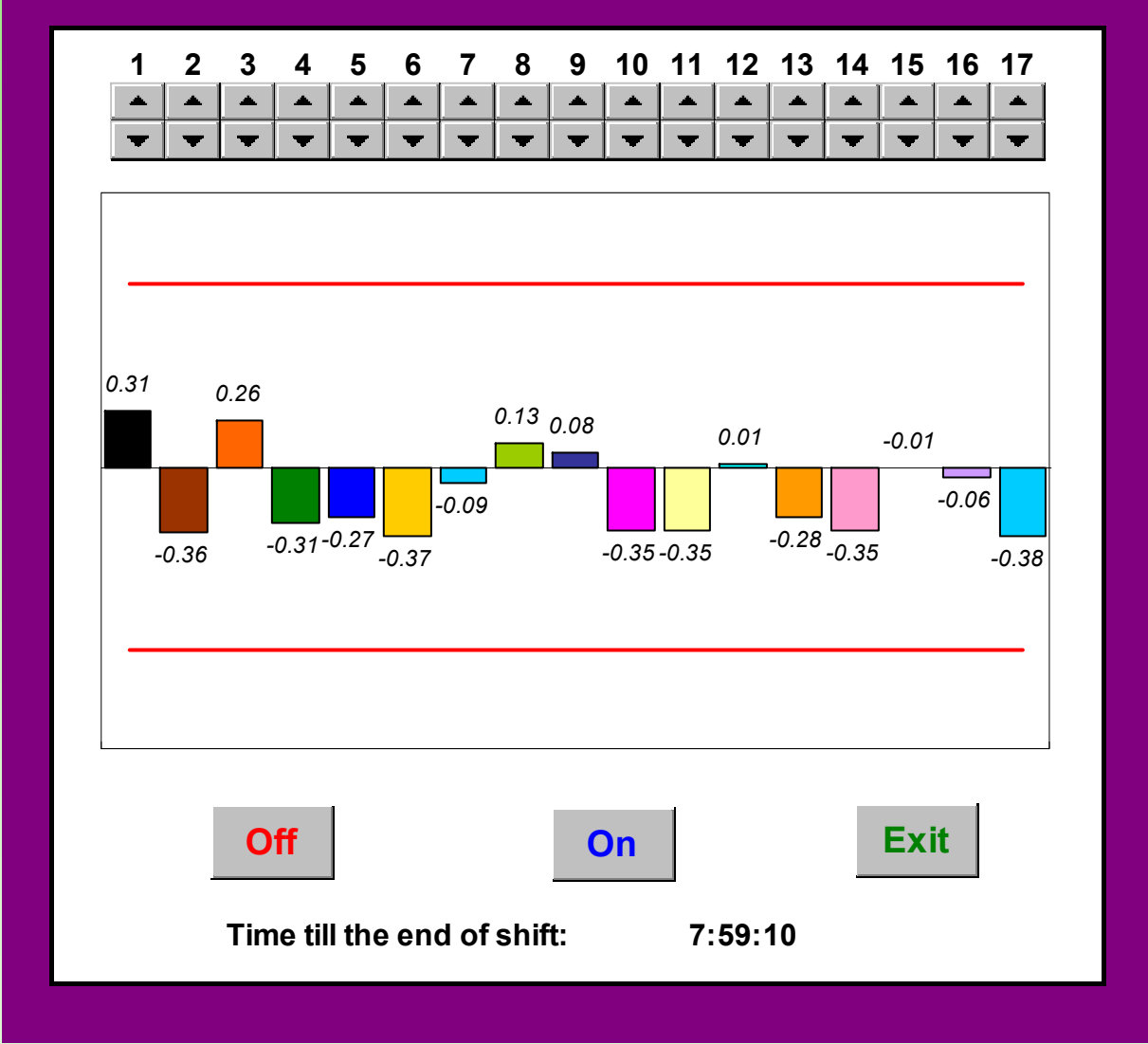
Пример данных о процессе

Реализации процесса s1, s2, ..., s54	Ключевые переменные процесса (датчики) X1, X2, ... , X17										
	X1	X2	X3	X4	X5	X6	X7	X8	X9	...	X17
s1	-1.19E-01	7.28E-01	-2.15E-02	5.22E-01	7.06E-04	7.32E-01	3.10E-04	-6.13E-04	-5.92E-05		9.74E-03
s2	-1.37E-01	7.28E-01	-2.89E-02	6.08E-01	7.09E-04	7.02E-01	6.58E-04	-1.22E-03	-1.49E-04		1.01E-02
s3	2.51E-02	-9.15E-02	6.73E-03	-1.13E-01	-9.07E-05	-7.58E-02	-2.29E-04	4.10E-04	5.65E-05		-1.43E-03
s4	-1.14E-01	6.70E-01	-2.18E-02	5.04E-01	6.50E-04	6.65E-01	3.83E-04	-7.34E-04	-7.96E-05		9.07E-03
s5	-7.93E-02	4.14E-01	-1.69E-02	3.51E-01	4.04E-04	3.98E-01	3.96E-04	-7.35E-04	-9.05E-05		5.78E-03
s6	1.51E-02	-6.38E-02	3.74E-03	-6.75E-02	-6.28E-05	-5.67E-02	-1.15E-04	2.07E-04	2.78E-05		-9.49E-04
s7	7.44E-02	-5.24E-01	1.11E-02	-3.24E-01	-5.06E-04	-5.45E-01	-1.73E-05	7.92E-05	-1.07E-05		-6.79E-03
s8	3.65E-02	-2.66E-01	5.12E-03	-1.59E-01	-2.56E-04	-2.78E-01	1.43E-05	-3.95E-07	-1.14E-05		-3.42E-03
s9	1.36E-01	-7.06E-01	2.89E-02	-6.01E-01	-6.88E-04	-6.77E-01	-6.83E-04	1.26E-03	1.56E-04	• • •	-9.86E-03
s10	-2.74E-02	3.60E-01	1.82E-03	1.12E-01	3.42E-04	4.12E-01	-4.31E-04	7.24E-04	1.22E-04		4.18E-03
s11	7.47E-02	-3.31E-01	1.80E-02	-3.34E-01	-3.25E-04	-2.99E-01	-5.30E-04	9.62E-04	1.28E-04		-4.84E-03
s12	-1.17E-01	7.02E-01	-2.16E-02	5.13E-01	6.81E-04	7.03E-01	3.40E-04	-6.63E-04	-6.76E-05		9.44E-03
s13	1.06E-01	-2.82E-01	3.23E-02	-4.82E-01	-2.85E-04	-1.87E-01	-1.25E-03	2.21E-03	3.14E-04		-4.99E-03
s14	7.39E-02	-5.28E-01	1.07E-02	-3.21E-01	-5.09E-04	-5.50E-01	2.49E-06	4.48E-05	-1.59E-05		-6.81E-03
s15	-9.87E-03	1.02E-01	-3.21E-04	4.17E-02	9.75E-05	1.13E-01	-8.29E-05	1.36E-04	2.44E-05		1.23E-03
s16	-1.06E-01	7.68E-01	-1.52E-02	4.62E-01	7.41E-04	8.03E-01	-2.54E-05	-2.68E-05	2.88E-05		9.90E-03
s17	-4.76E-02	2.66E-01	-9.52E-03	2.10E-01	2.59E-04	2.61E-01	1.92E-04	-3.61E-04	-4.19E-05		3.65E-03
					• • •						
s54	6.61E-02	-5.40E-01	7.19E-03	-2.85E-01	-5.19E-04	-5.78E-01	1.81E-04	-2.67E-04	-6.23E-05		-6.78E-03

Карты Шухарта (1927)

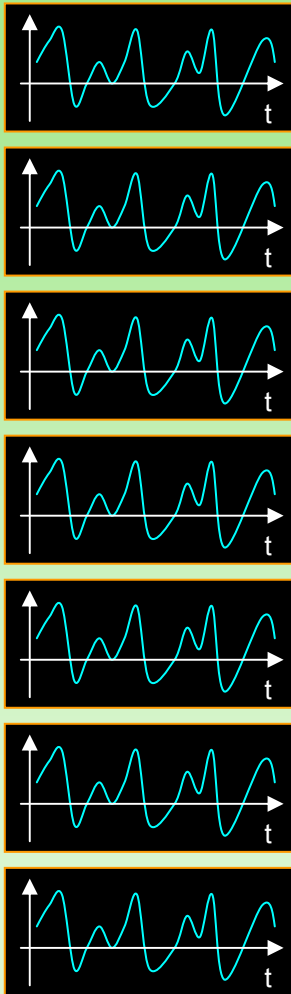


Контроль процесса (просто игра)

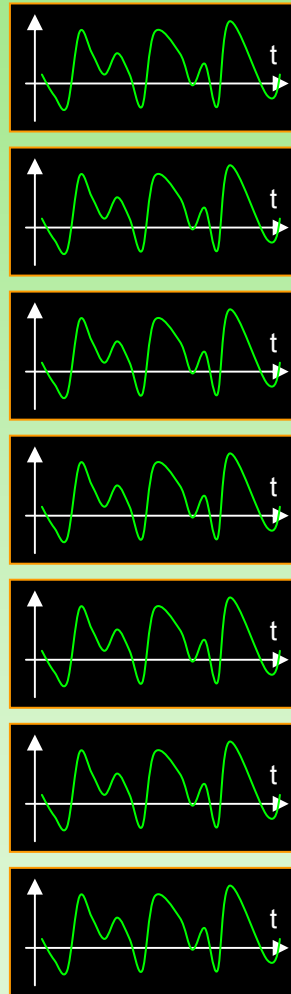


Контроль с помощью SPC

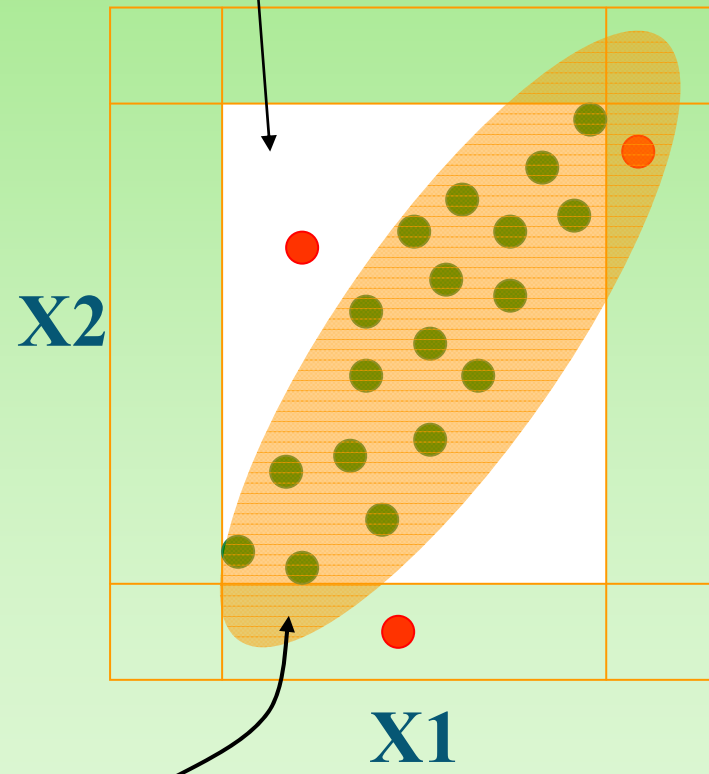
X1



X2



Одномерная область контроля



Многомерная область контроля

Многомерный статистический контроль процессов (MSPC)

Цель

Контроль хода процесса

Концепция

Анализ накопленных исторических данных о процессе

Методы

Методы анализа многомерных данных (PCA, PCR, PLS)

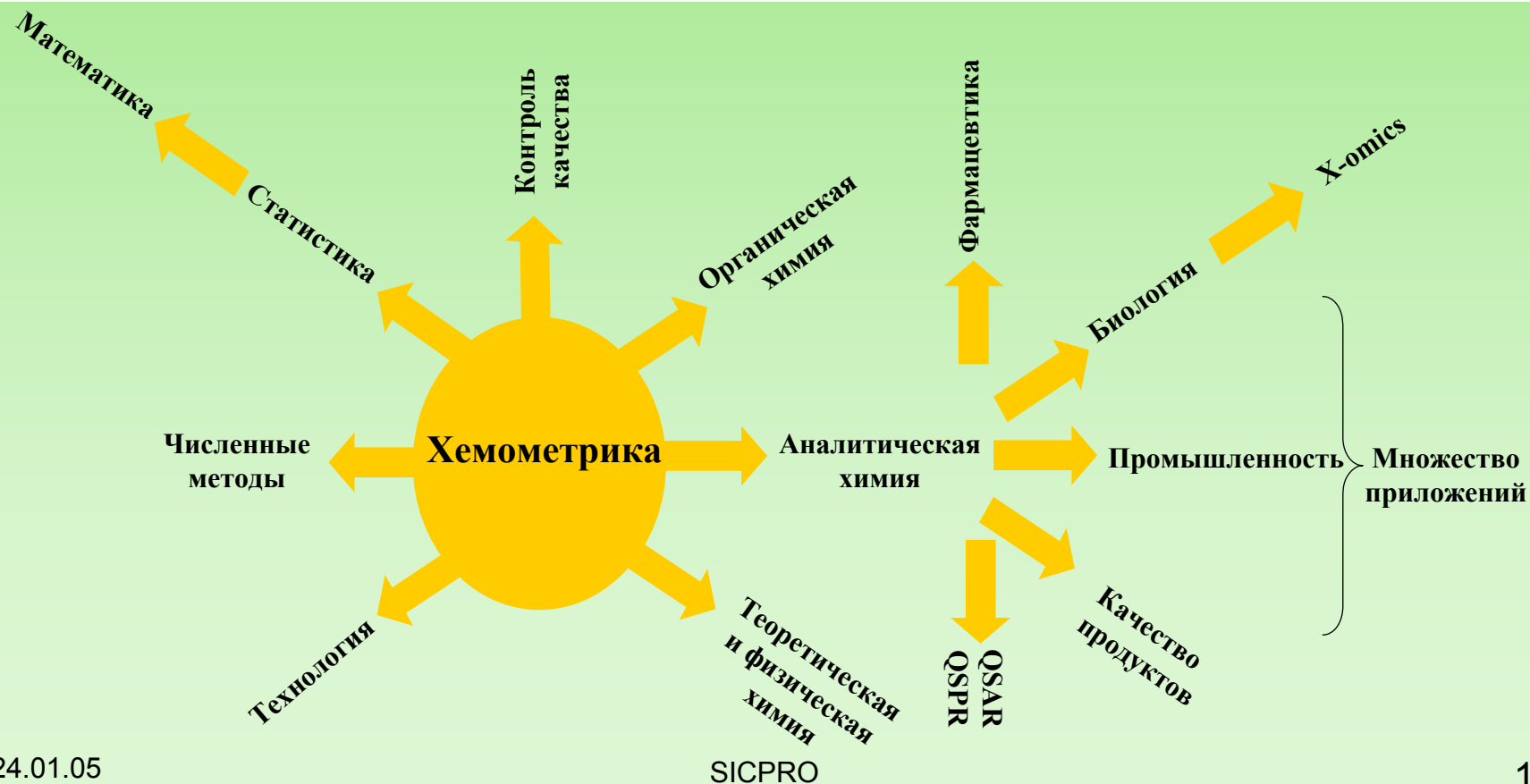
Реализация

Построение графиков счетов и нагрузок, т.е. *хеометрика*

Хемометрика

Хемометрика - это научная дисциплина, находящаяся на стыке химии и математики, предметом которой являются математические методы исследования химических данных

сайт PXO <http://rsc.chph.ras.ru>



Что такое хемометрика?

- Хемометрика имеет дело с **данными** (зачастую с очень большими), поэтому хемометрика - это подраздел информатики (Data mining)
- Данные, которые исследует хемометрика по большей части происходят из **химии**, поэтому хемометрика - это подраздел химии (Analytical chemistry)
- Методы, которые использует хемометрика ориентированы на **формальное** моделирование (Soft modeling)

Почему «хемо-» ?

- Хемометрика родилась из задачи анализа химических **спектров**
- Спектроскопия – наилучший метод получения информации по ходу процесса (**on-line**) в режиме реального времени: быстро и без влияния на процесс
- «Хемо» подчеркивает **практическую**, а не статистическую значимость применяемых методов

Почему «-метрика» ?

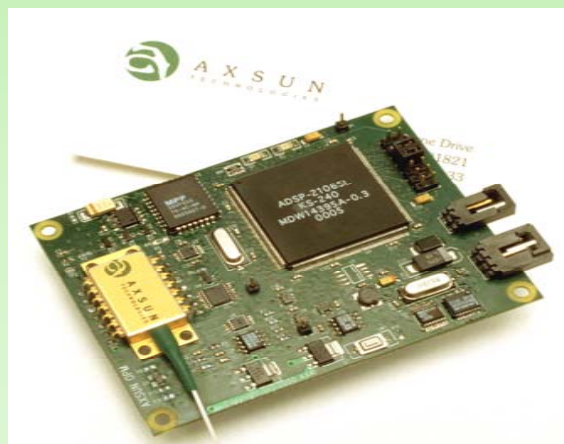
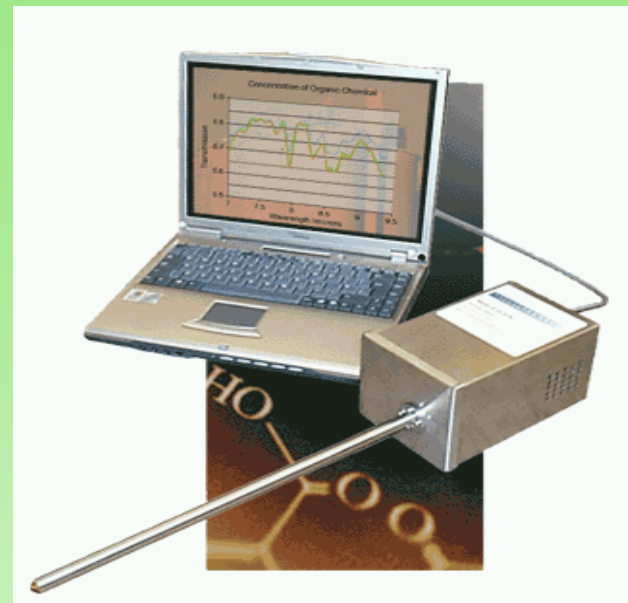
- Хемометрические методы легко и плодотворно переносятся в другие области, например, в психологию, биологию, геологию, и т. д.
- Хемометрика активно эксплуатирует математику статистику, линейную алгебру
- ‘It is easier to teach a chemist statistics than to teach chemistry to a statistician.’ (Svante Wold)

Зачем в MSPC нужна хемометрика?

Потому, что все больше данных о процессах получают с помощью современных, эффективных приборов

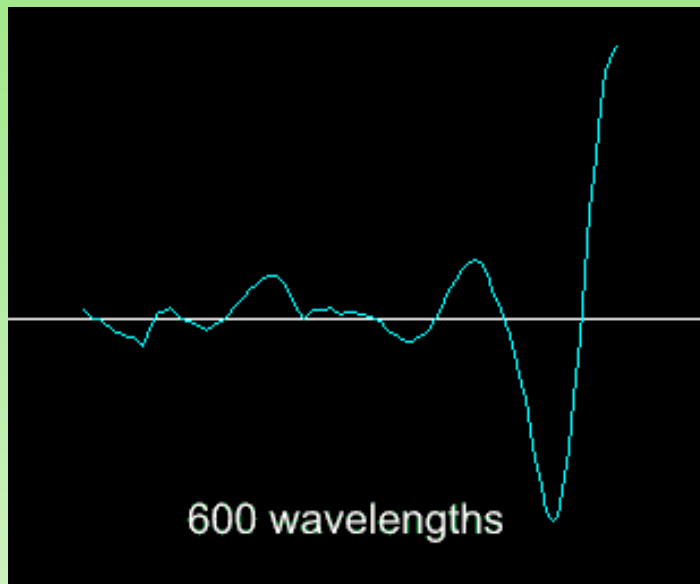
- Макропеременные (температура, pH, давление, ...)
- Спектроскопия (УФ, ИК, БИК, ...)
- Хроматография (ЖХ, ГХ, ...)
- Гибридные методы (ЖХ/МС, ЖХ/ЯМР, ...)
- Видео образы и гиперспектры (2D, 3D)

Эволюция спектрофотометров

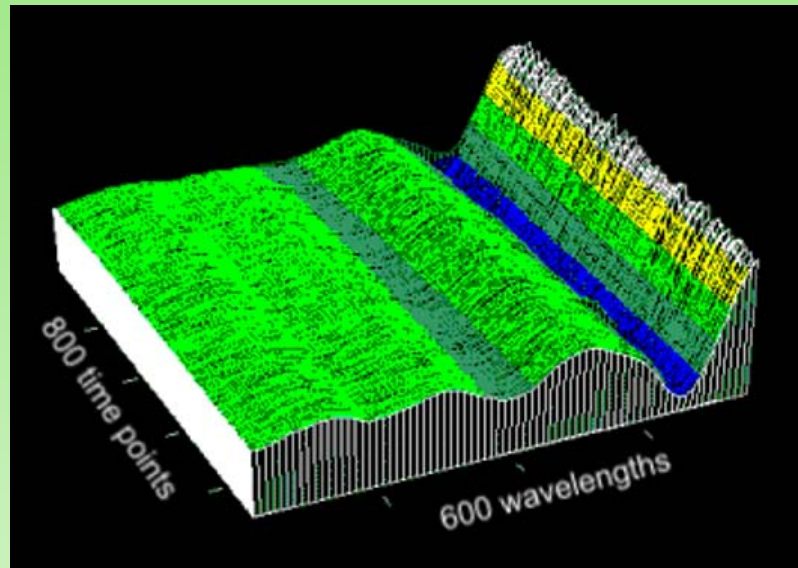


Много переменных и много измерений

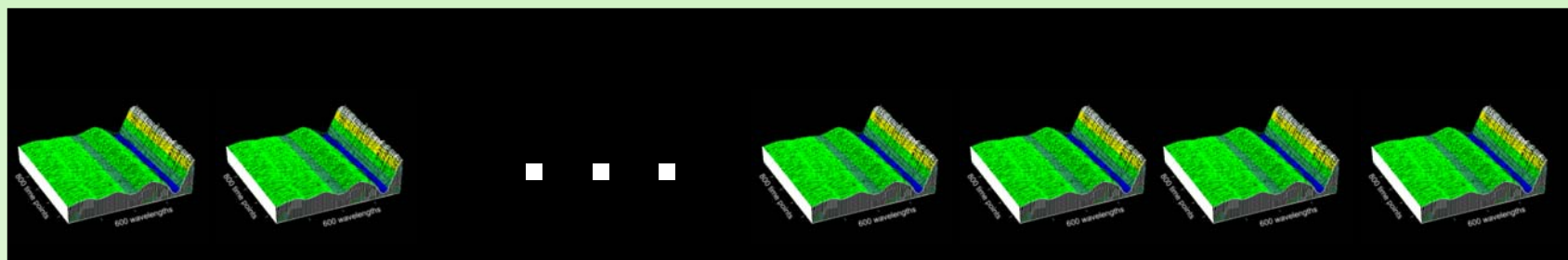
Одно измерение – спектр (600 точек)



Один цикл – 800 спектров (времен)



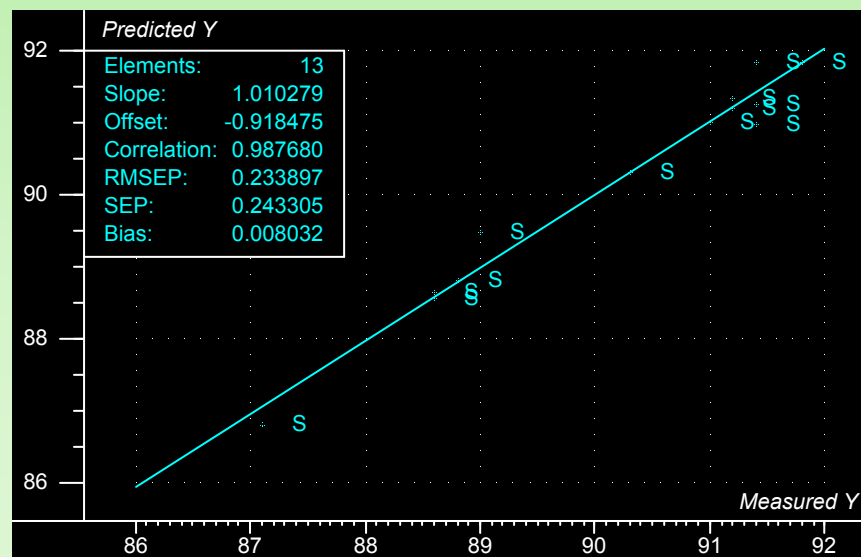
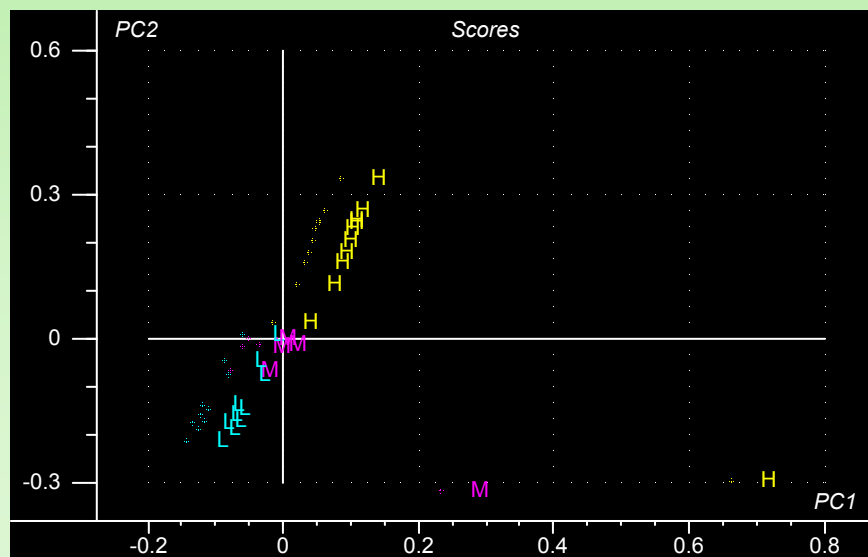
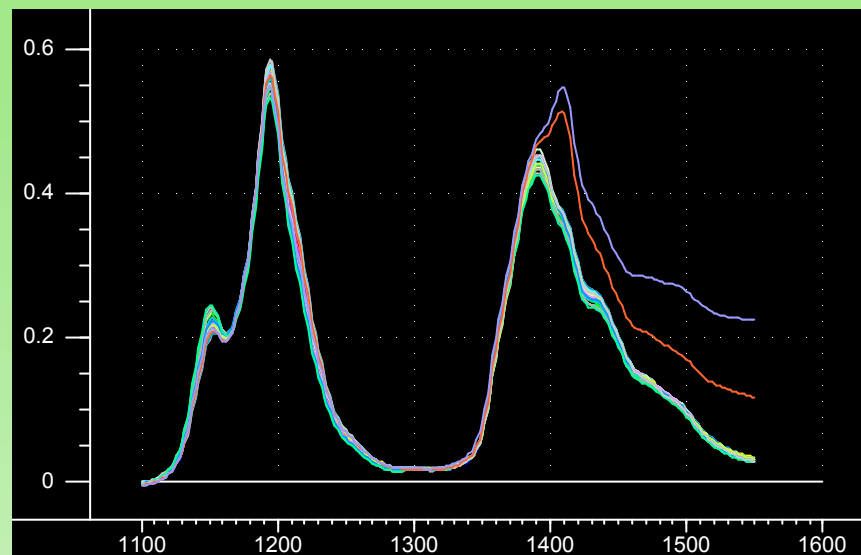
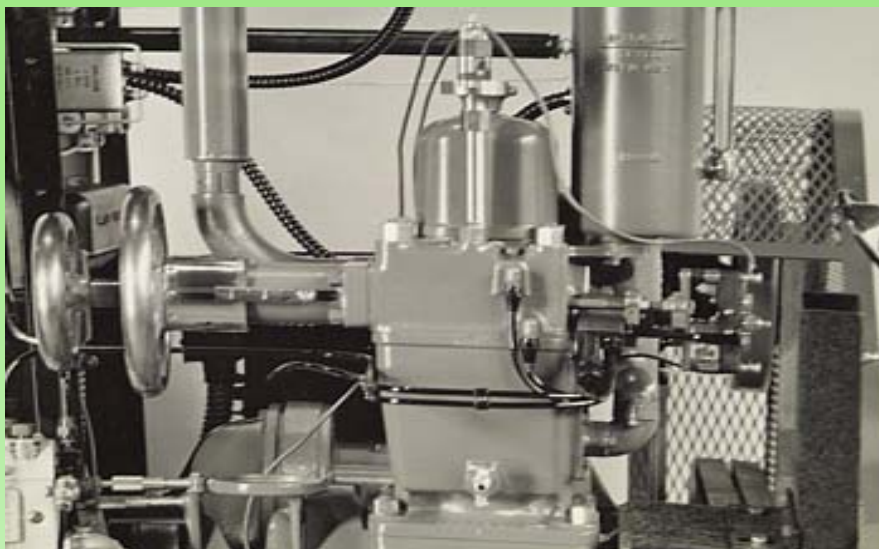
Один массив данных – 200 образцов (циклов)



Формальные и содержательные модели

	Содержательные “Hard” models	Формальные “Soft” models
Откуда	Физика, химия,	Из данных
Формула	$y=f(\mathbf{x}, \mathbf{a})+\varepsilon$	$y=\mathbf{X}\mathbf{a}+\varepsilon$
Параметры	Имеют физ. смысл	Физически бессмысленны
Проблемы	Построить модель	Обработать данные
Назначение	Экстраполяция	Интерполяция
Пример	Хим. кинетика	ANOVA

Определение качества бензина по ИК-спектру



Основная задача хемометрики

Заменить **прямые** измерения, которые либо –

- невозможны
- дороги
- длительны

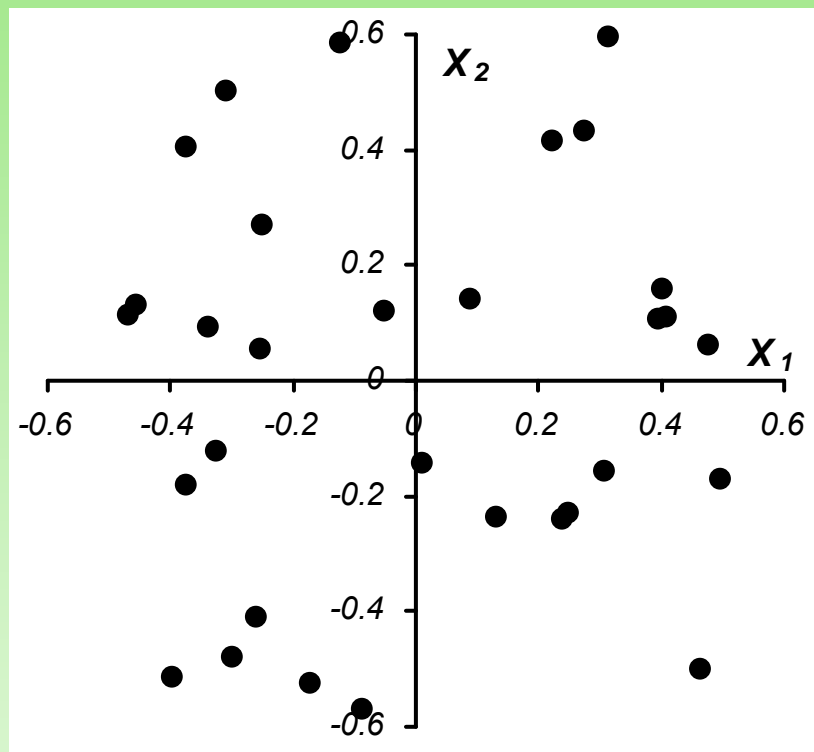
на **косвенные** измерения, которые –

- доступны
- дешевы
- быстры

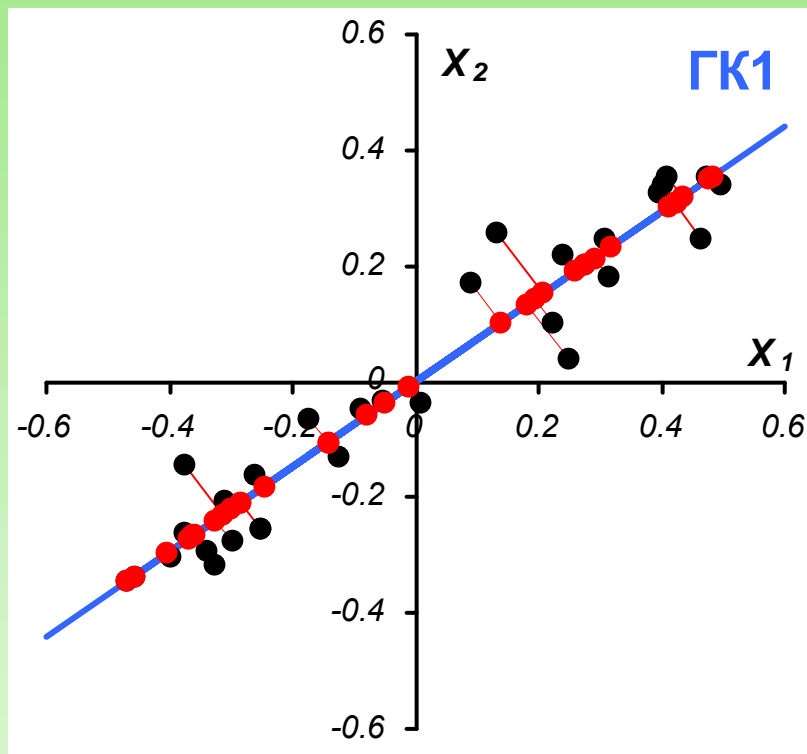
с последующей их **обработкой** (калибровкой).

Главные направления и проекции

Данные без структуры



Данные со скрытой структурой



$$X_2 = aX_1 + E$$

Проекция на подпространство

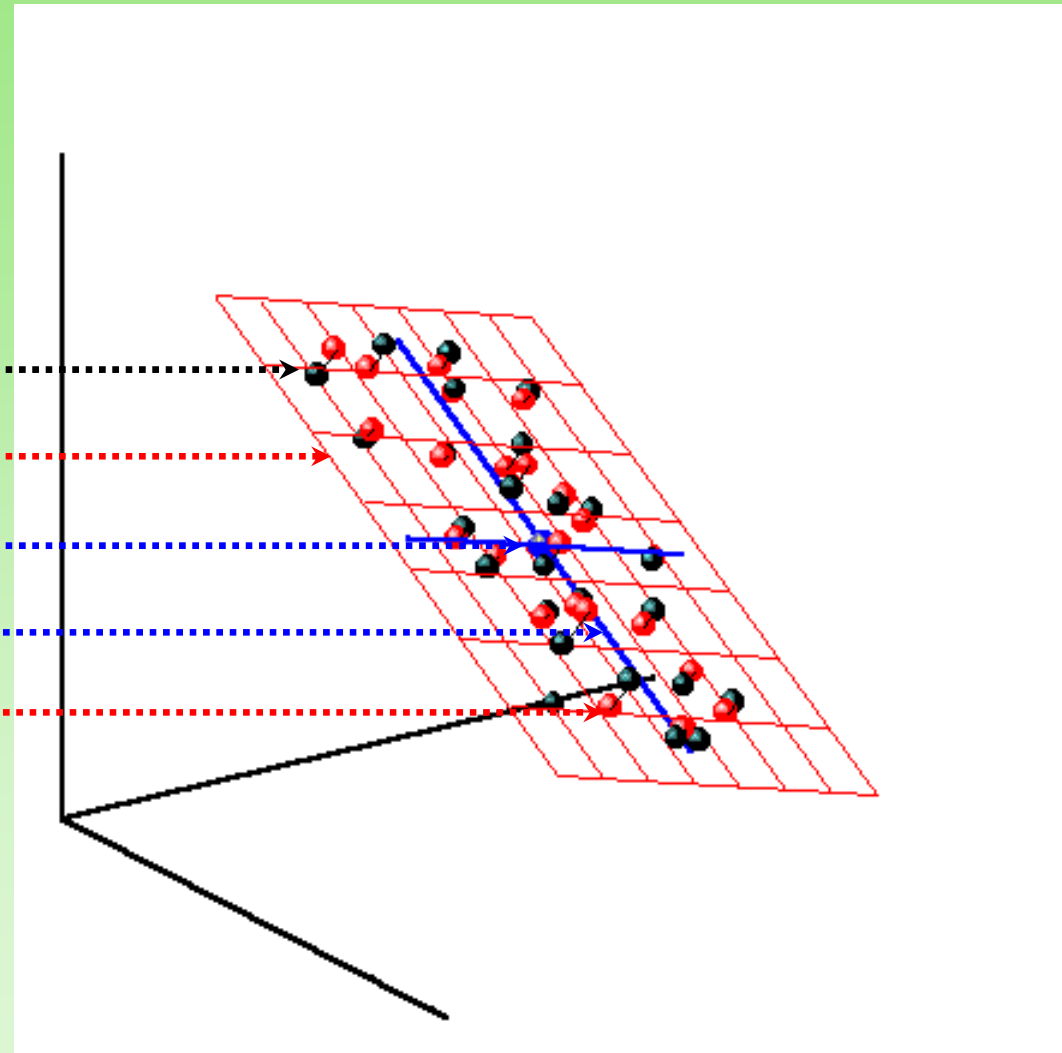
Исходные данные

Подпространство

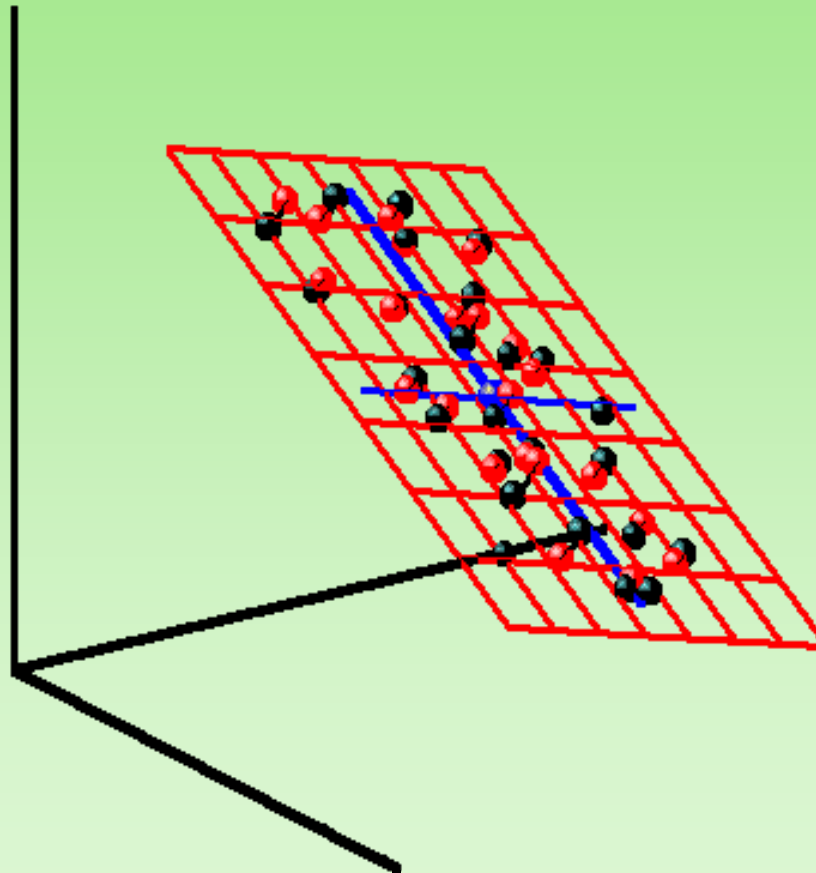
Центр данных

Главные компоненты

Проекции данных

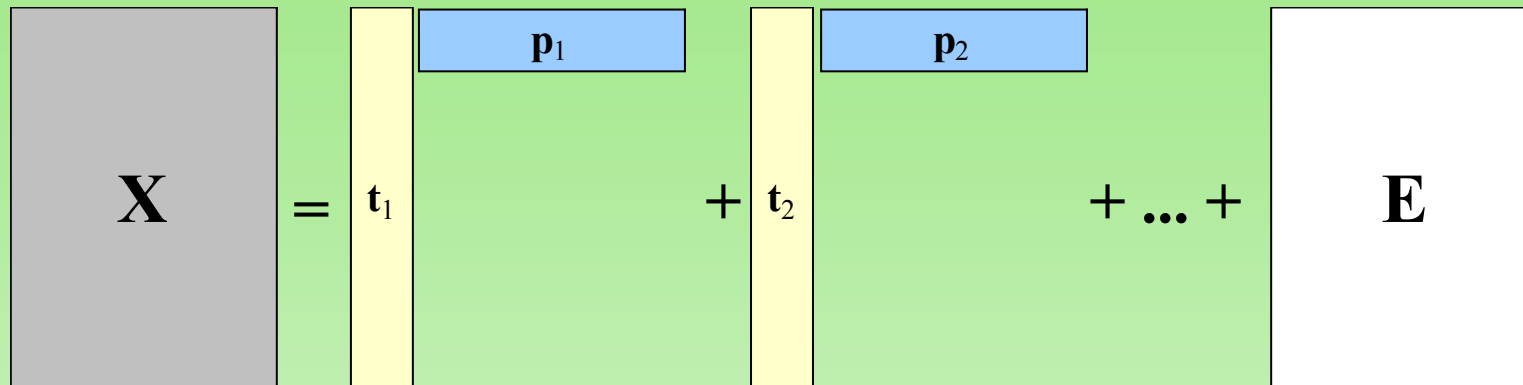


Представление данных в подпространстве



Проекционные методы: PCA

Один блок данных. Метод главных компонент (МГК)



$$\mathbf{t} = \mathbf{X}\mathbf{w} \Leftarrow \max |\mathbf{X}\mathbf{w}|^2 \text{ при условии } |\mathbf{w}|=1 \Leftrightarrow \mathbf{X}^t \mathbf{X}\mathbf{w} = \lambda \mathbf{w}$$

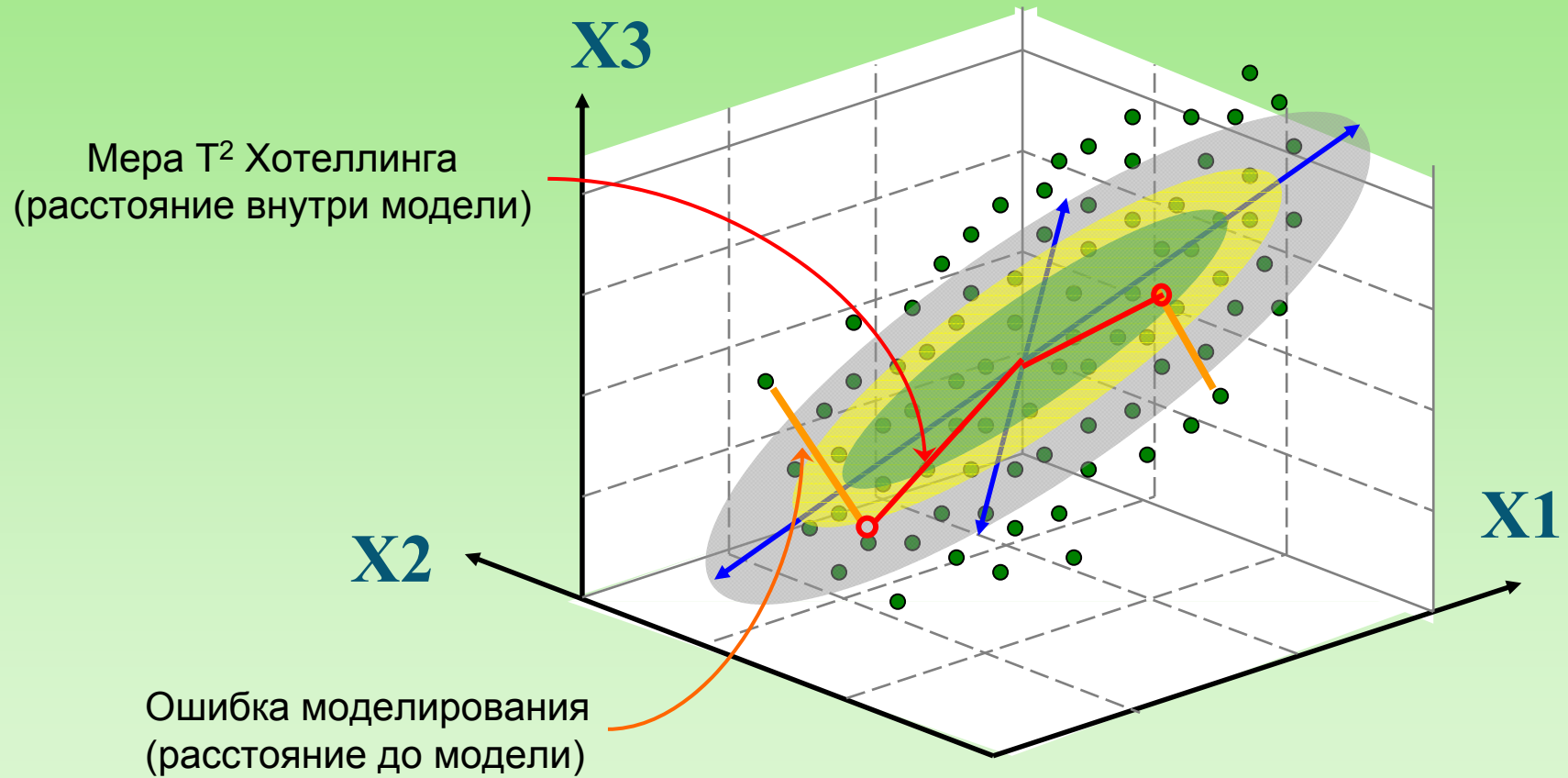
\mathbf{X} - матрица данных, \mathbf{E} - матрица ошибок, обе $(n \times p)$

\mathbf{T} - матрица счетов: $(n \times k)$, \mathbf{P} - матрица нагрузок: $(k \times p)$

k - число главных компонент ($k \ll p$)

Karl Pearson, 1901

Контроль с помощью MSPC



Графики MSPC (Пример процесса)

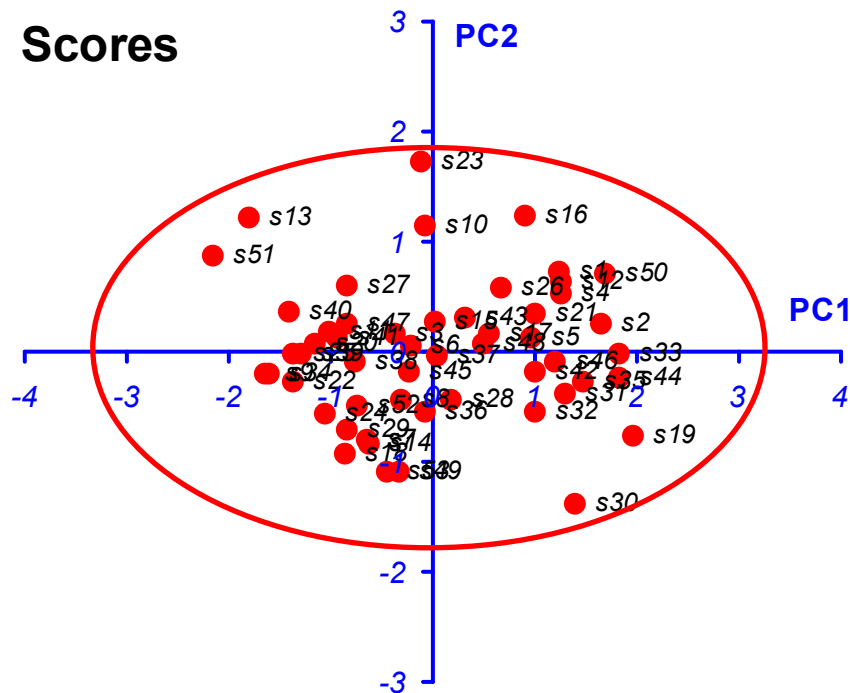
Образцы

Счета Т

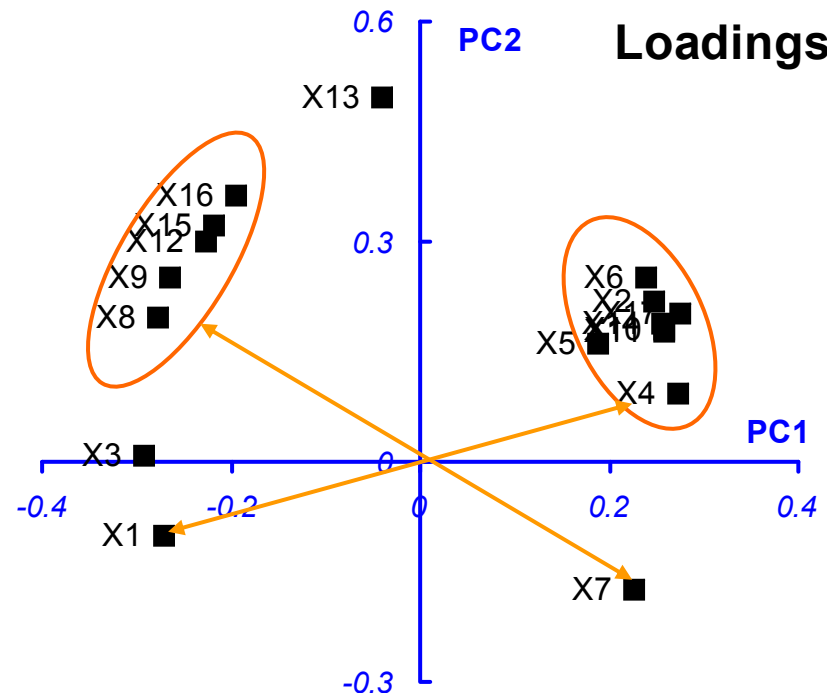
Переменные

Нагрузки Р

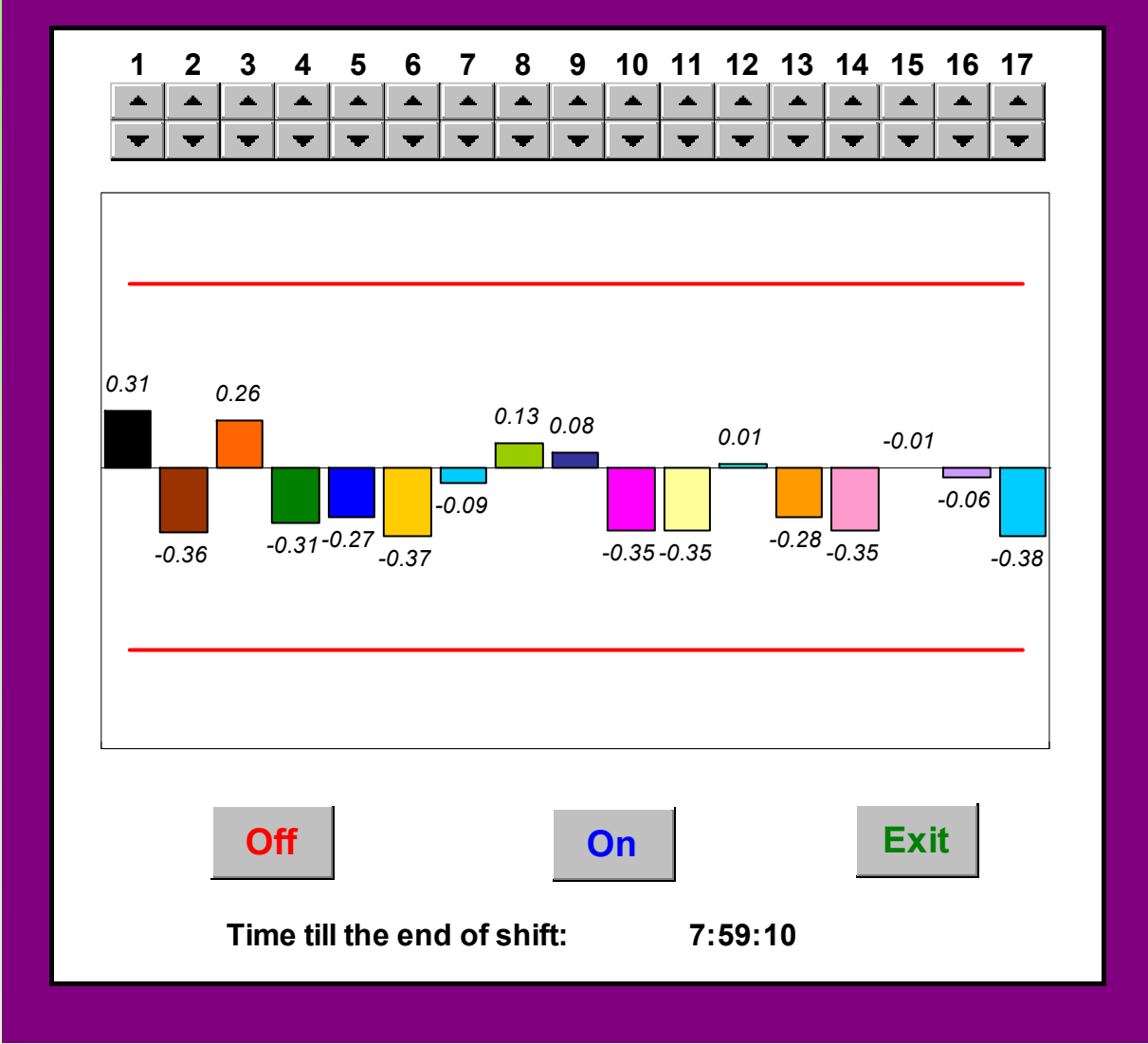
Scores



Loadings



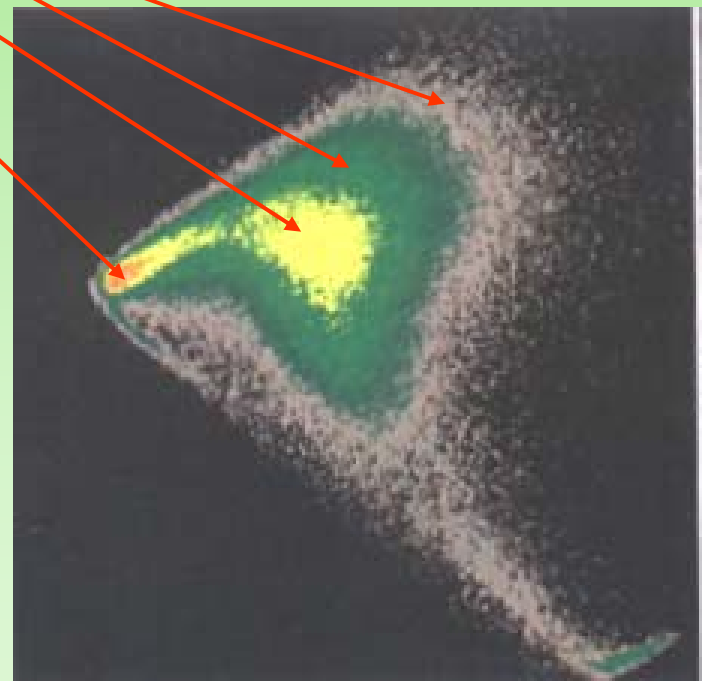
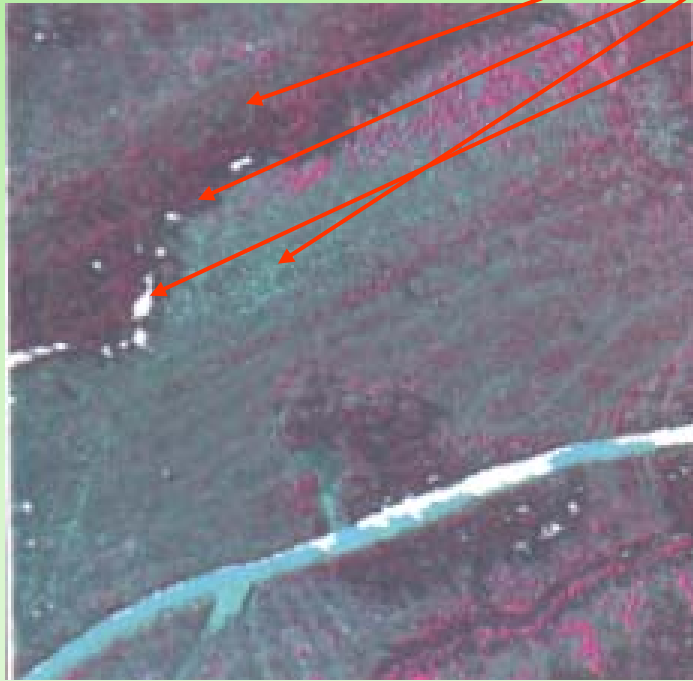
Контроль процесса (не просто игра)



Исследование состояния лесов (Канада) с помощью анализа изображений (MIA)

By Paul Geladi, Multivariate Image Analysis, Wiley & Sons, 1996

Исходный аэроснимок отражения Старые деревья Состав в темн Он же в пространстве ГК



Периодические (batch) процессы

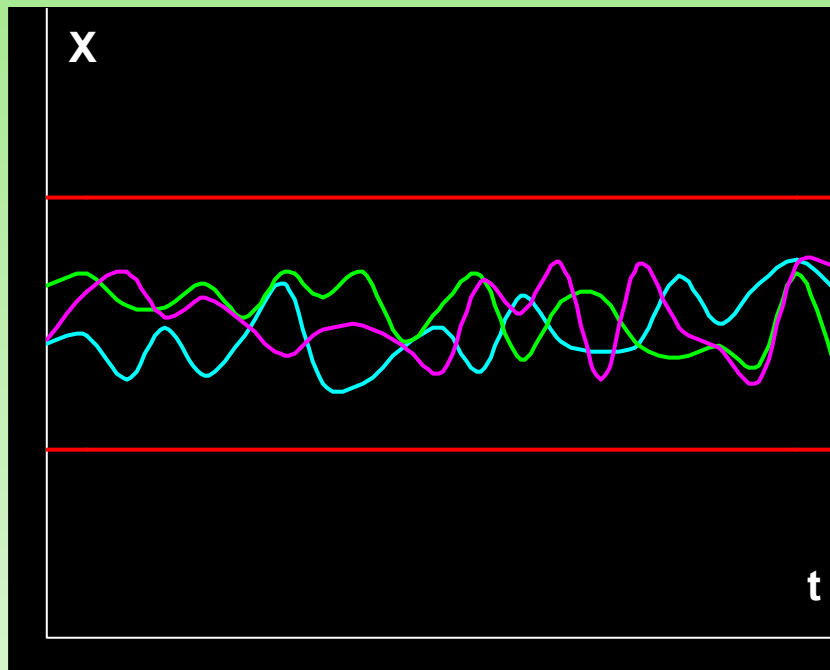
Периодические процессы применяются при производстве лекарств, полимеров, пищевых продуктов, и т. п. Они характеризуются –

- Конечной продолжительностью
- Кинетикой протекания
- Изменчивостью, как внутри, так и между циклами
- Частой сменой производимых продуктов

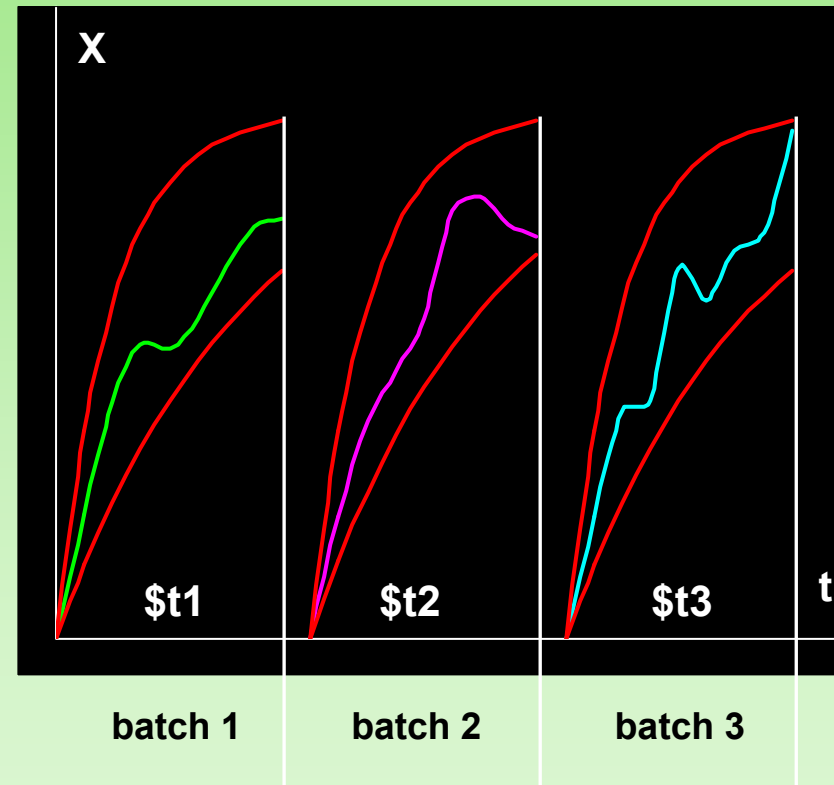
В непрерывных процессах важны только взаимоотношения между переменными, тогда как в периодических важно знать как переменные **меняются во времени**. Поэтому здесь особенно важен контроль в реальном времени.

Сравнение двух типов процессов

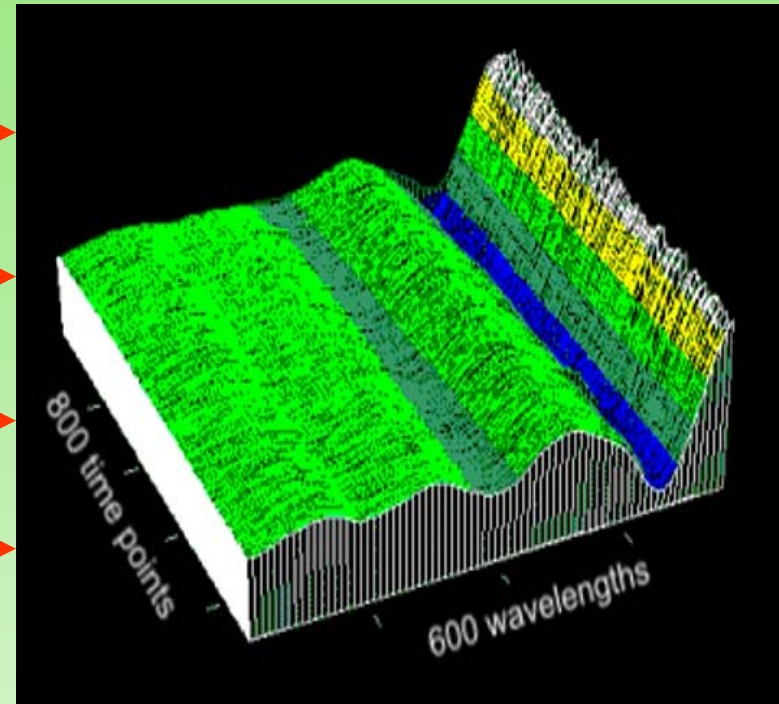
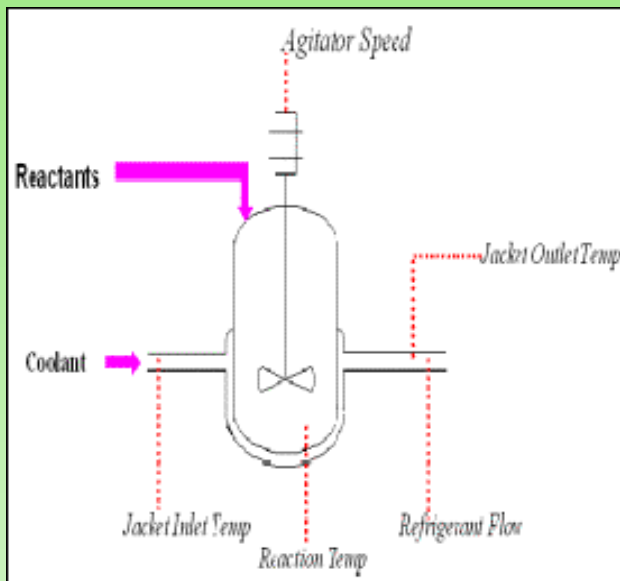
Непрерывный процесс



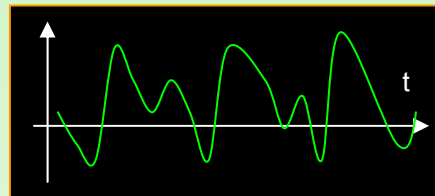
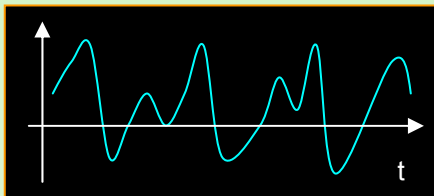
Периодический процесс



Контроль в фармацевтике

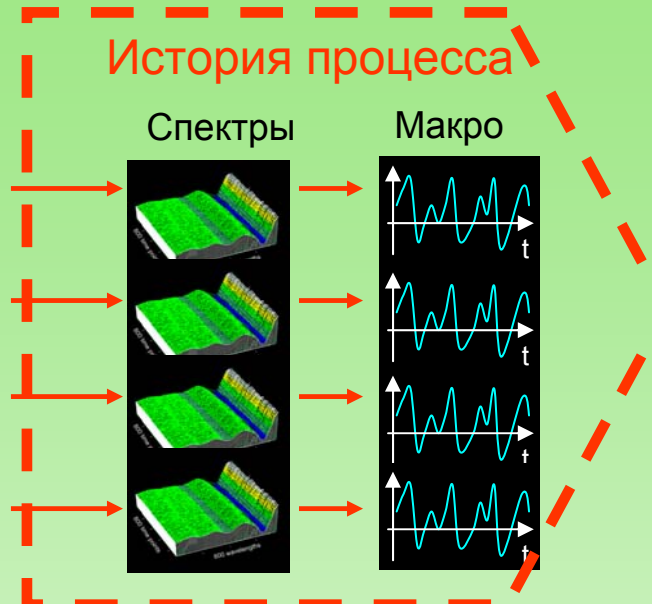
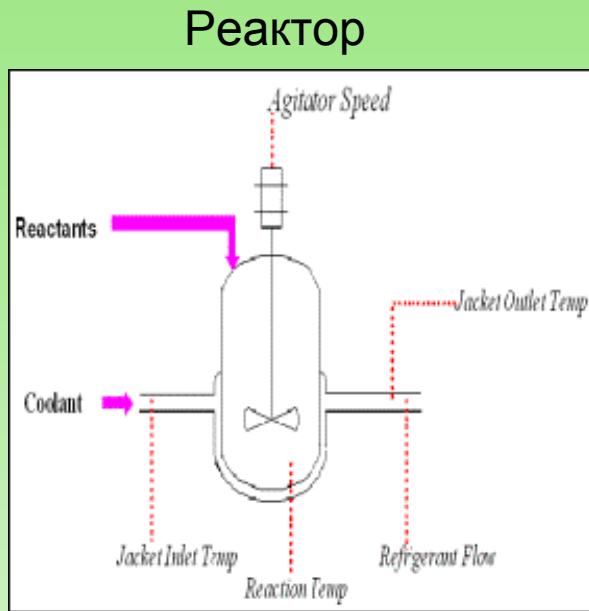


Макропараметры: температура, давление

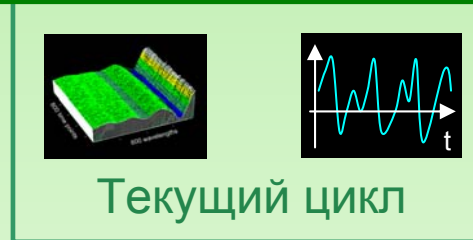
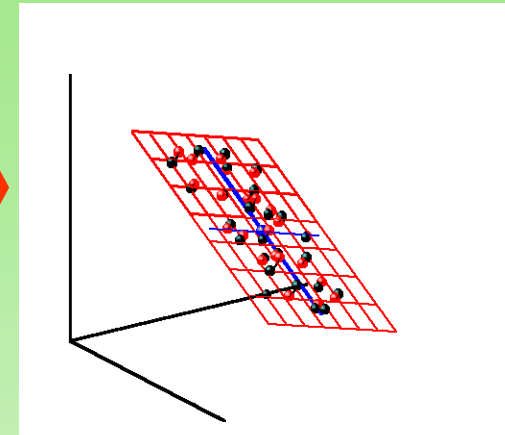


$$\begin{aligned} \frac{dA}{dt} &= -k_1 A; & A(0) &= A_0 \\ \frac{dB}{dt} &= k_1 A - k_2 B; & B(0) &= B_0 \\ \frac{dC}{dt} &= k_2 B; & C(0) &= C_0 \end{aligned}$$

MSPC в фармацевтике



Проекционная модель



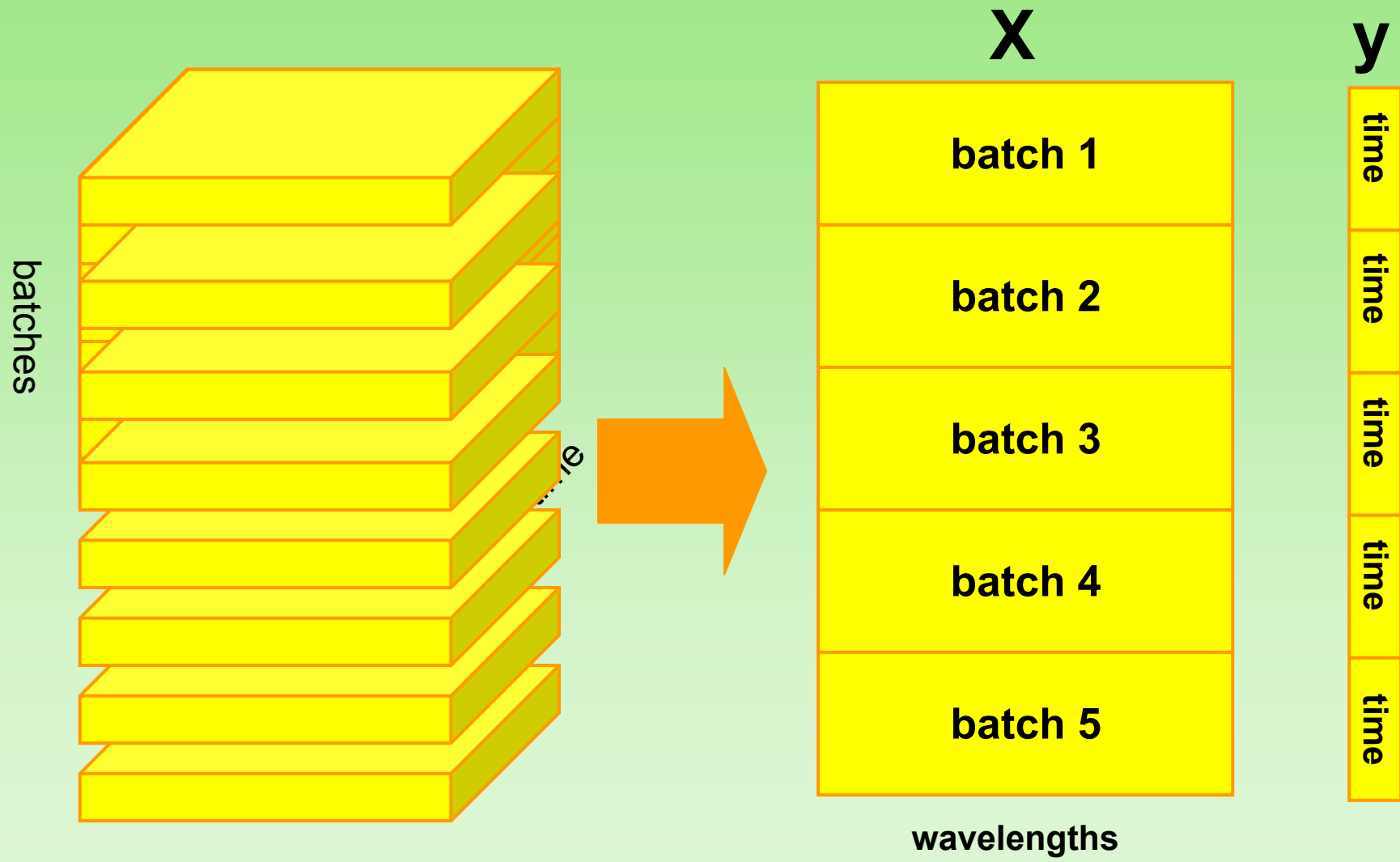
MSPC Контроль

BothSets.sip - Predicted scores for [1]

BothSets.sip - Contribution Scores
Batch: 01p7-025 - average batch at 200:10:03 16:15:00 (pre-maturity 5), weight 1

Обратная связь

Трехмодальные (3-way) данные



Регрессия

$$X \times a = y + \varepsilon$$

$$p \gg n$$

$$\text{rank}(X) < p$$

Проекционные методы 2: PCR, PLS

Два блока данных. Метод проекций на латентные структуры (ПЛС)

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad \mathbf{Y} = \mathbf{UQ}^t + \mathbf{F}$$

$$\mathbf{t} = \mathbf{Xw} \leftarrow \max |\mathbf{Y}^t \mathbf{Xw}|^2 \text{ при условии } |\mathbf{w}| = 1 \Leftrightarrow \mathbf{X}^t \mathbf{Y} \mathbf{Y}^t \mathbf{X} \mathbf{w} = \lambda \mathbf{w}$$

\mathbf{X} - матрица данных, \mathbf{E} - матрица X -ошибок, обе $(n \times p)$

\mathbf{Y} - матрица откликов, \mathbf{F} - матрица Y -ошибок, обе $(n \times m)$

\mathbf{T} - матрица X -счетов: $(n \times k)$, \mathbf{P} - X -матрица нагрузок: $(p \times k)$

\mathbf{U} - матрица Y -счетов: $(m \times k)$, \mathbf{Q} - Y -матрица нагрузок: $(m \times k)$

k - число главных компонент ($k \ll p$)

Herman Wold, 1973 & Agnar Höskuldsson, 1988

Вполне реальный пример

Real Time Quality Control (Batch monitoring) of Pharmaceutical Production Processes.

Chris Ambrozic and John Parsons
Umetrics Inc., Kinnelon, NJ 07405

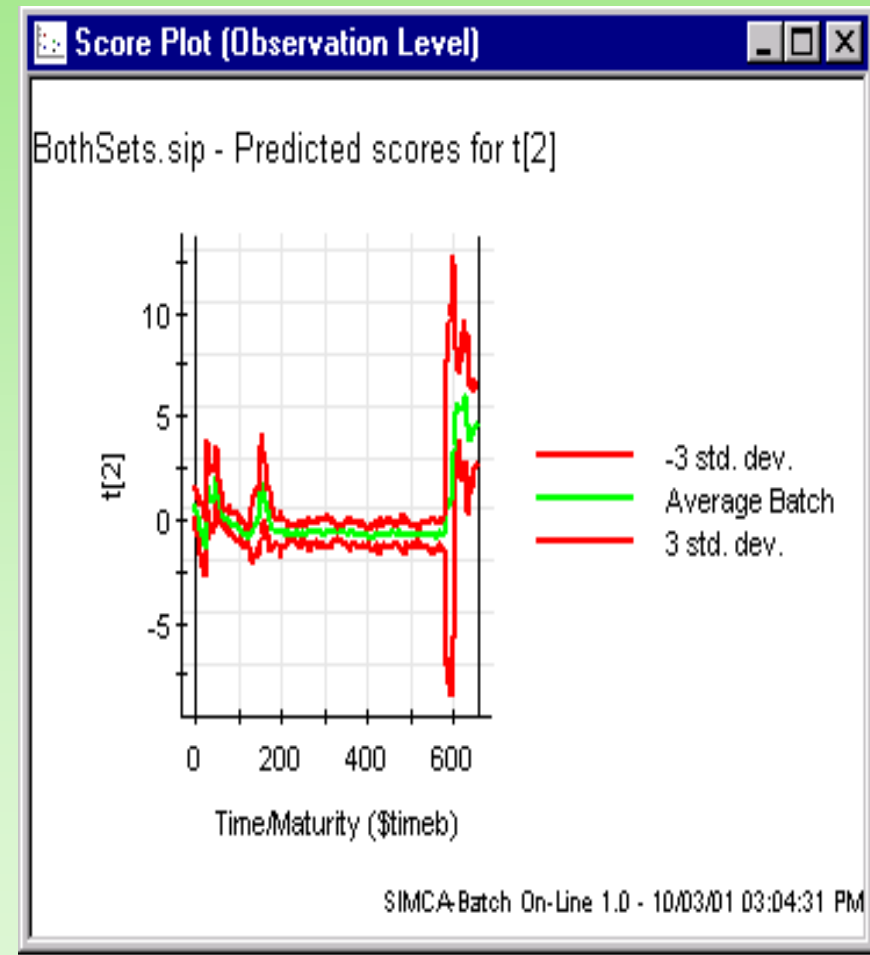
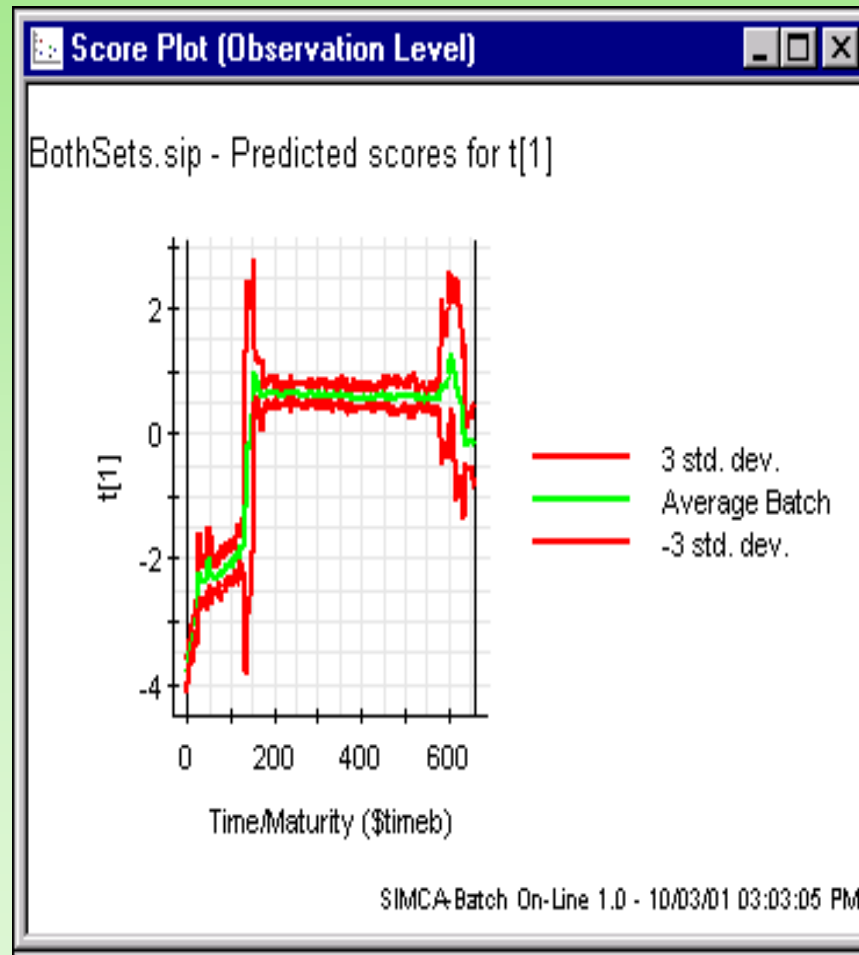
www.umetrics.com

Presented at FDA, 28 Nov., 2001



MSPC Контроль по первым двум ГК

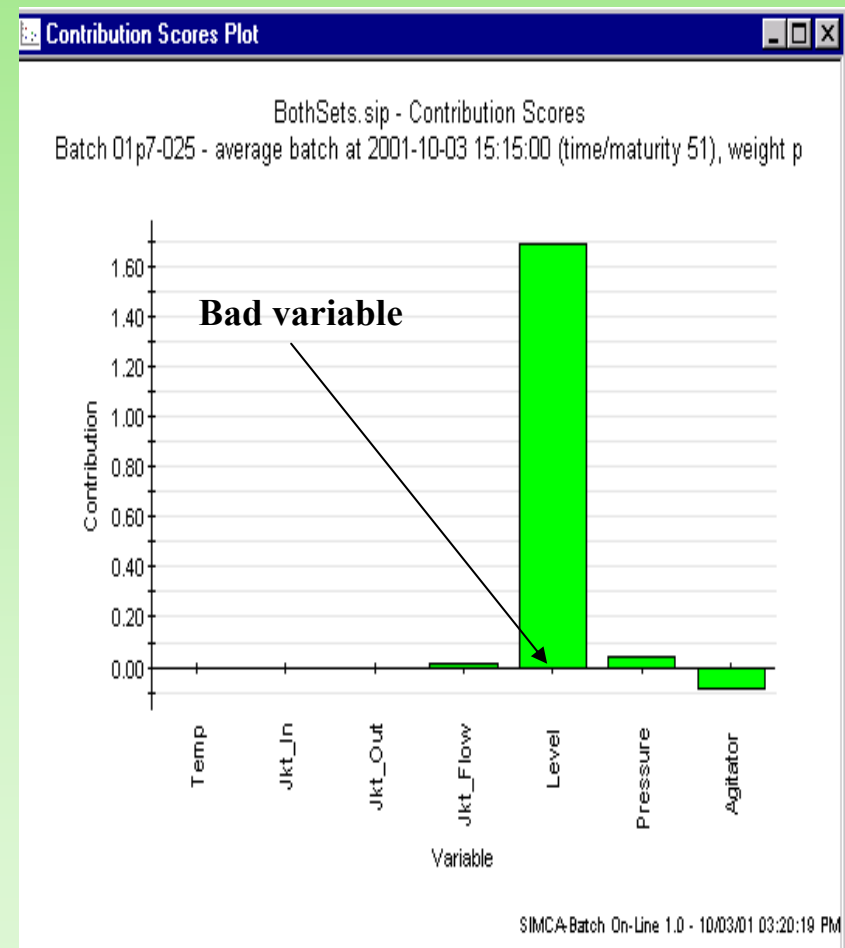
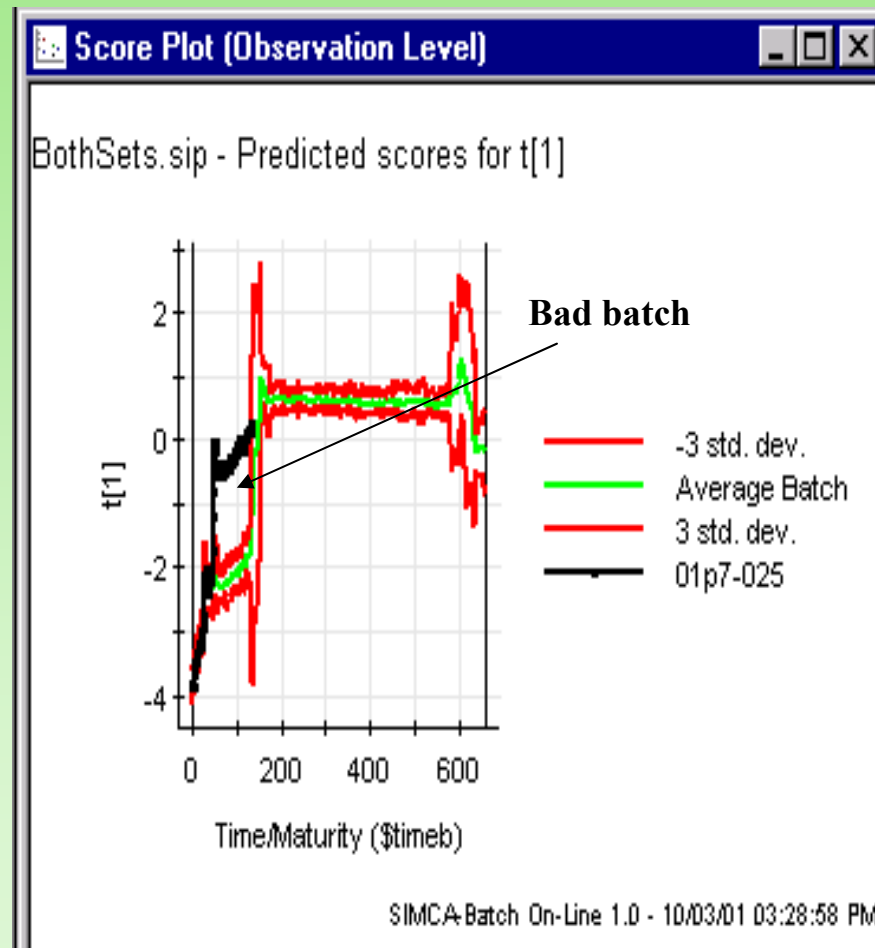
Средняя траектория хороших циклов (зеленая) $\pm 3\sigma$ (красные).
Хорошие циклы должны лежать в этих пределах.



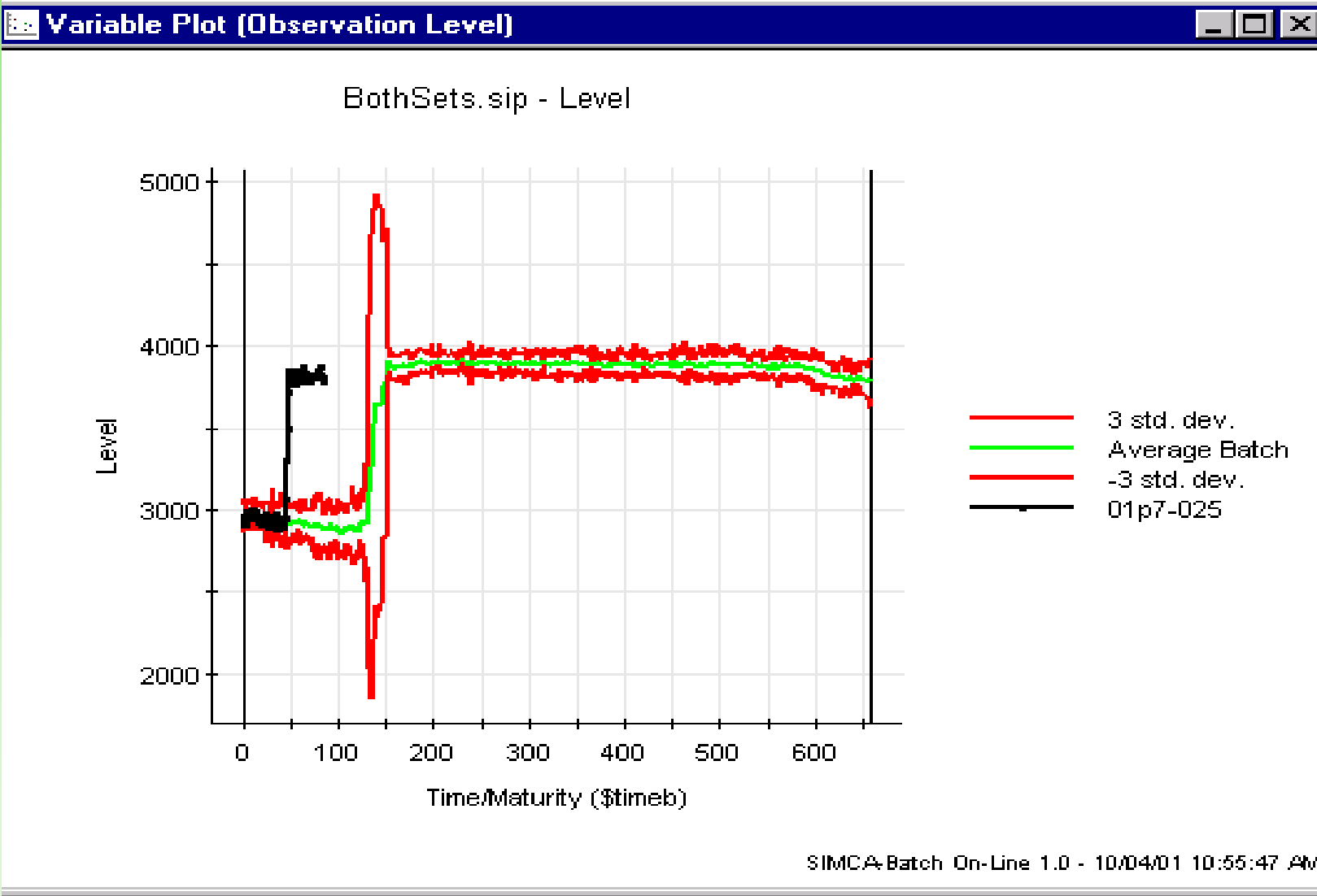
Контроль процесса в реальном времени

Плохой процесс (черный) выходит за пределы допустимых отклонений по ГК1.

Причина выясняется с помощью графика вкладов переменных в первую ГК.



Контрольная карта для этой переменной



Выводы

1. MSPC заслуживает самого пристального внимания, особенно после **решения FDA** от сентября 2004 г.
2. MSPC использует простые **корреляционные зависимости**, а не сложные содержательные модели.
3. MSPC в сочетании с хемометрикой может обеспечить контроль сложных биохимических процессов в **реальном времени**
4. MSPC уже 20 лет используется во всем мире, но в России практически **не известен**, хотя именно здесь он мог бы дать максимальный эффект.

Литература о проекционных методах



Благодарности

Richard Brereton, University of Bristol, UK

Kim Esbensen, Ålborg University, Denmark

Svante Wold & Nouna Kettaneh-Wold, Umetrics, Sweden

Julian Morris, University of Newcastle, UK

Karl Booksh, Arizona State University, USA

Karl Ring, Computas AS, Norway

Paul Geladi, Biomass Technology and Chemistry, Sweden

