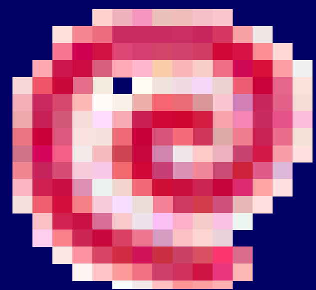
The Power of Multivariate Statistical Process Control

Susan Albin
Rutgers University

ICIEOM, Fortaleza Brazil
October 2006

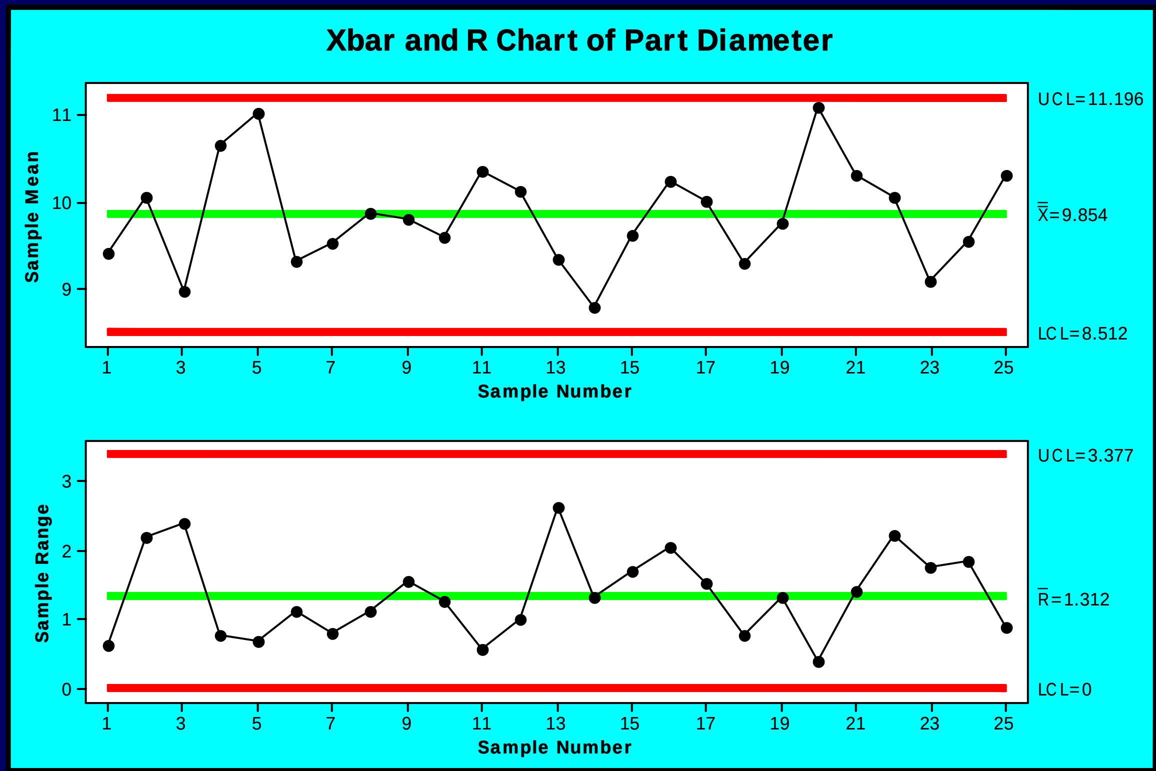
Outline of Talk

- Intro to statistical process control
- Multivariate statistical process control - How do we handle so many variables at the same time?
- The power of MSPC
  - – Start-up in medical devices
  - – Selecting process settings in steel manufacturing



Statistical Process Control
One Variable

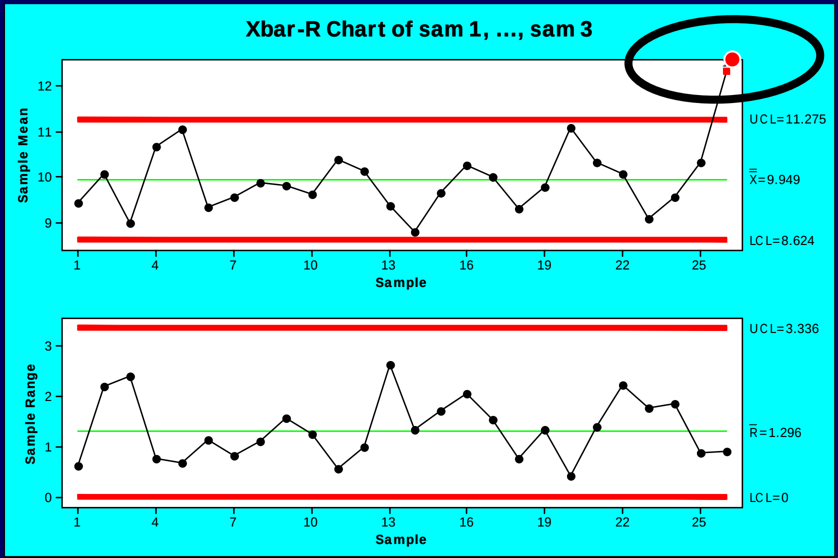
- Monitor process or product variable - SPC Charts
- Shewhart 1920s
- Sample every hour
- Plot sample average and range



Statistical Process Control: Two Phases

- **I MODEL**:
  - Collect baseline data during good production
  - Confirm stationary, find mean and sd

- **II MONITOR**:
  - Stop production points fall outside limits



Basic Idea of Process Control

Model

Characterize Baseline
Product & Process Data

Monitor

Monitor current process data
Is it consistent with baseline?

If not, take action

What if there are Multiple Product & Process Variables?

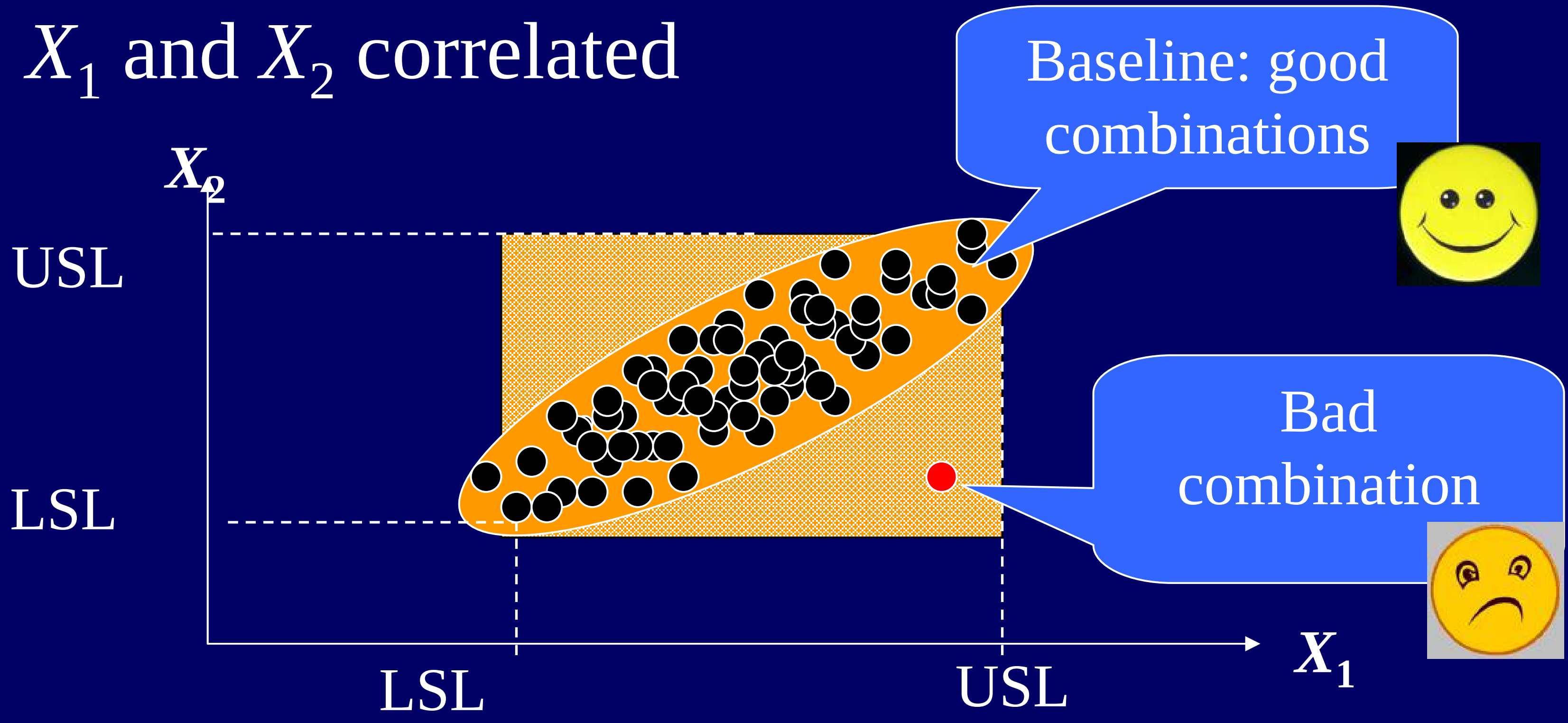
- Process variables (100s)
  - temperatures, pressures, speeds
  - collected by sensors, stored in IT systems
  - Excellent for monitoring

- Product variables (10s)
  - diameter, tensile strength
  - available but with delay

Variables Correlated

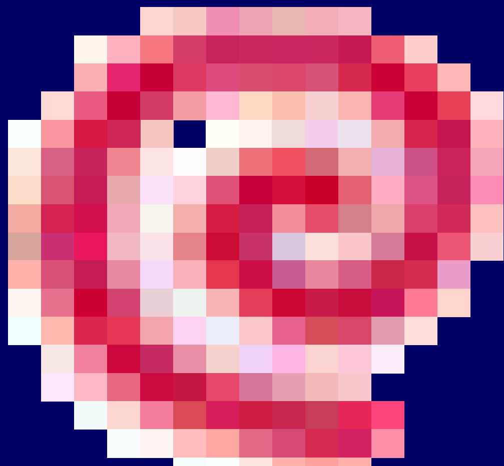
Why not monitor Process Variables One-at-a-Time?

- X_1 and X_2 correlated



Power of Multivariate Process Control

Model and monitor the <u>relationships</u> among variables



What Does Baseline Data Look Like?

$$
\begin{array}{c}
obs\,1 \\ obs\,2 \\ \vdots \\ obs\,n
\end{array}
\begin{bmatrix}
x_{1,1} & \cdots\cdots & x_{1,100} & y_{1,1} & \cdots & y_{1,3} \\
x_{2,1} & \cdots\cdots & x_{2,100} & y_{2,1} & \cdots & y_{2,9} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{n,1} & \cdots\cdots & x_{n,100} & y_{n,1} & \cdots & y_{n,9}
\end{bmatrix}
$$

Process variables X

Product variables Y

Multivariate SPC Requires Xs and Ys Synchronized

Commonly Available

	Time	Variable	Value
Process	4:00	Moisture in dough	…….
Process	4:00	Oven temp Chamber 1	…….
Process	4:00	Oven temp Chamber 6	…….
Product	4:00	Cookie height	…….

Necessary for MSPC

Time	Variable	Value
4:00	Moisture in dough	…….
4:21	Oven temp Chamber 1	…….
4:26	Oven temp Chamber 6	…….
4:32	Cookie height	…….

PLS, a Multivariate Statistical Model

- PLS
  - partial least squares regression
  - projection to latent structures
- Reduces dimensionality
- Emphasizes process variables that really affect product

What Does the PLS Model Look Like?

- PLS Components, Ts, are linear functions of process variables – coordinate translation

Just 3-5 PLS comps, Ts

$$T_1 = w_{11}X_1 + w_{12}X_2 + w_{13}X_3 \cdots$$

$$T_2 = w_{21}X_1 + w_{22}X_2 + w_{23}X_3 \cdots$$

$$T_3 = w_{31}X_1 + w_{32}X_2 + w_{33}X_3 \cdots$$

Many process variables, Xs

Select weights to emphasize process variables that are important to product outcomes

PLS Considers Both Process & Product Variables

- PLS Components, Ts
  - In terms of process variables, Xs
  - Collected frequently $\Rightarrow$ available

YET...

- Capture information about the *Y*s
  - PLS components capture correlation between *X*s and *Y*s

Construct PLS Components, *T*'s, with Optimization

$$T_1 = w_1X_1 + w_2X_2 + w_3X_3 + \cdots\cdots$$
$$U_1 = c_1Y_1 + c_2Y_2 + c_3Y_3 + \cdots\cdots$$

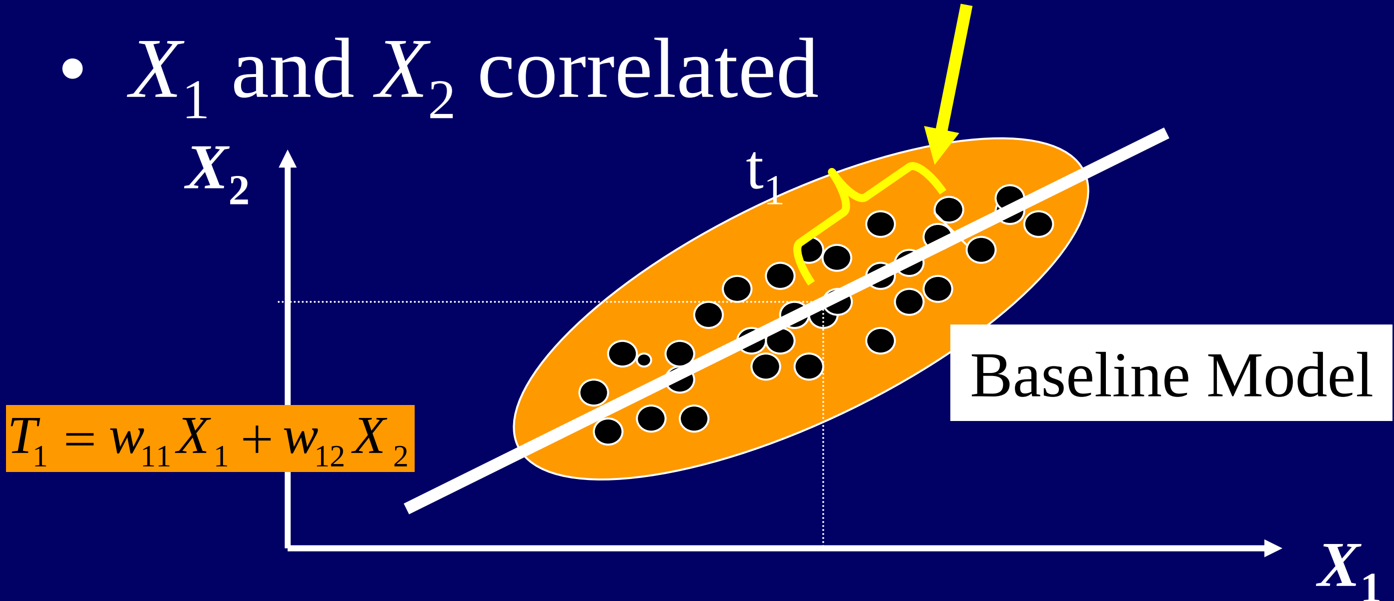
- **For T_1 solve for *w*'s and *c*'s (normalized):**
  Objective function: Max $Cov(T_1, U_1)$

- **For T_2 solve for *w*'s and *c*'s:**
  Objective function: Max $Cov(T_2, U_2)$
  $s.t.\ T_2 \perp T_1$

Nipals algorithm, SAS, S+, PLS Toolbox (with MATLAB), Umetrics - Simca

PLS Model for Two Variable Example

- X_1 and X_2 correlated



t_1: dist from center to Projection of point to line

- Model reduces data dimension: $(x_1,x_2) \Rightarrow t_1$
- Which line? maximize correlation between t_1 and product variables

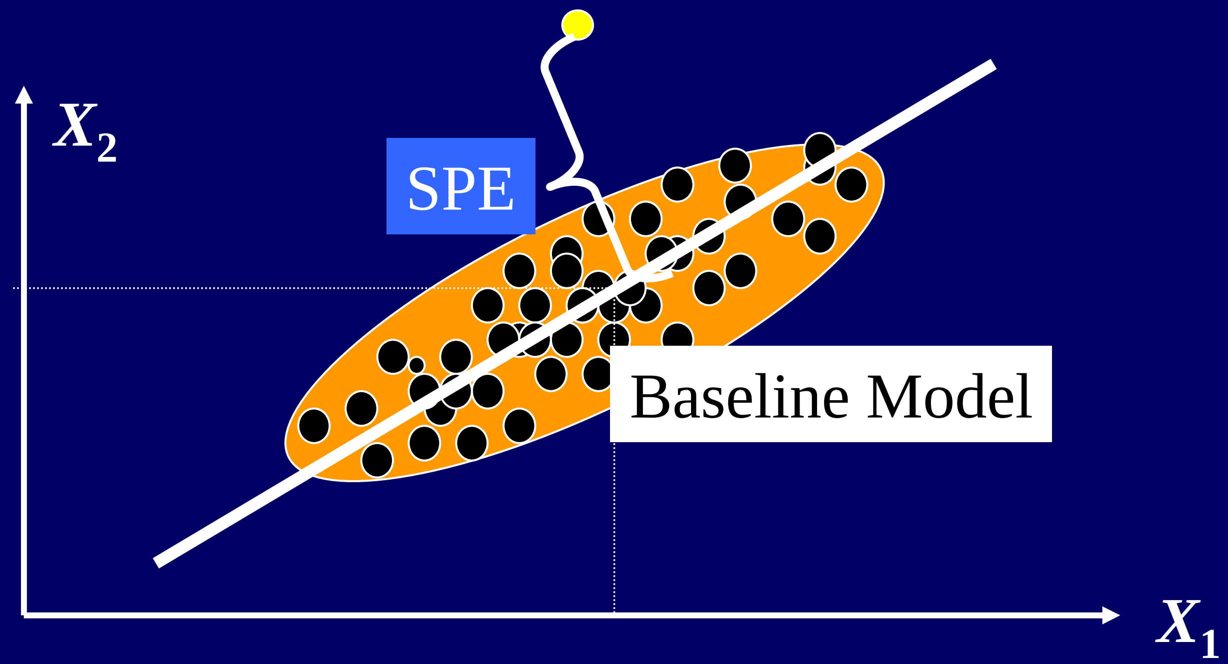
Monitoring: SPE Measures Distance Between New Point & Baseline

Squared Prediction Error, SPE

--SPE too big!

--Process vars not consistent with baseline

--Product vars may not be acceptable



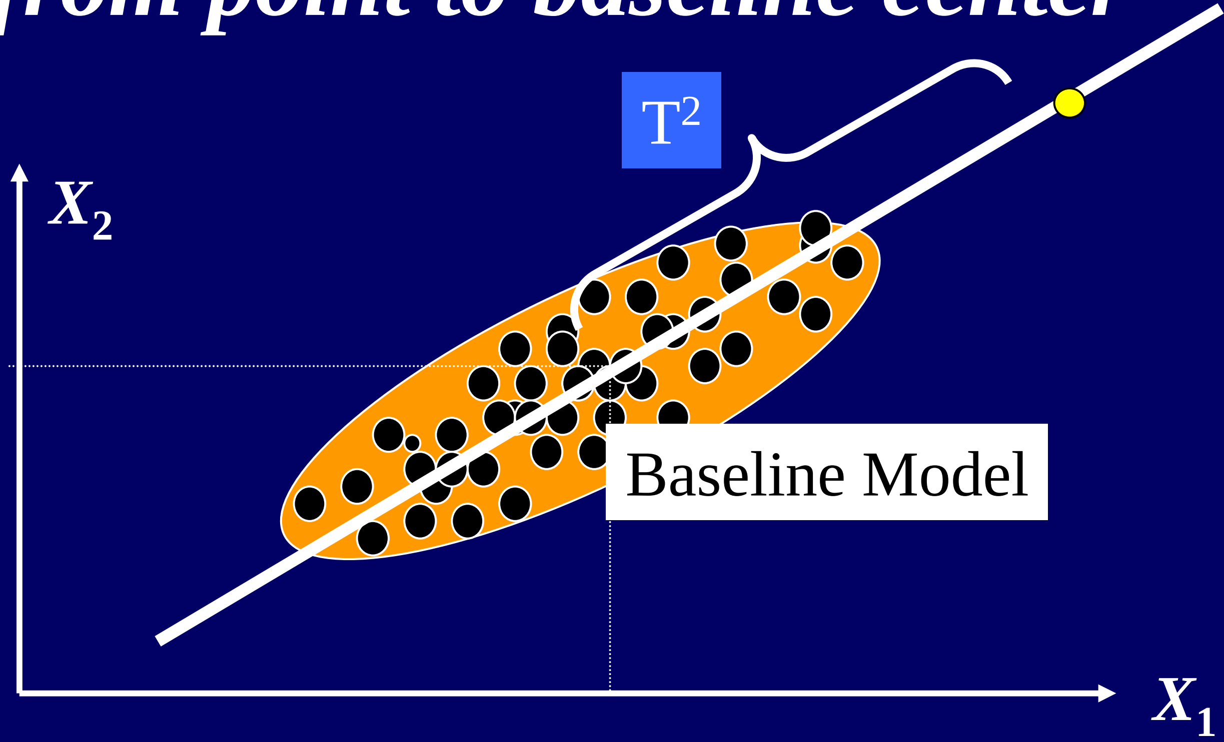
Monitoring: T^2 Measures Distance Between New Point & Baseline Center

T^2, Dist^2 from point to baseline center

--T^2 too big! (SPE=0)

--Process vars inconsistent with baseline

--Product vars may not be acceptable

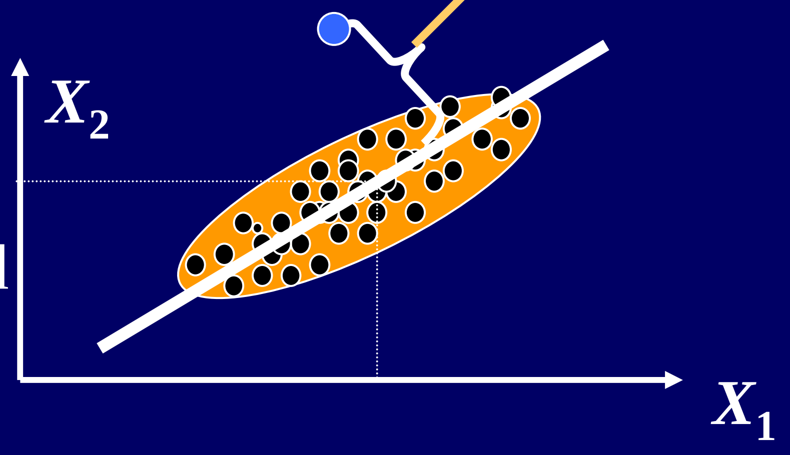


How to Calculate SPE

- Baseline Results
  - Weights – calculate Ts from Xs
  - Loadings – calculate Xs (many) from Ts (few)

- Production
  - observe x's & calculate t's
  - Find predicted Xs
  - compute SPE, distance between observed & predicted

$$SPE = \sum_{i=1}^{n} \left(x_i - \hat{x}_i \right)^2$$



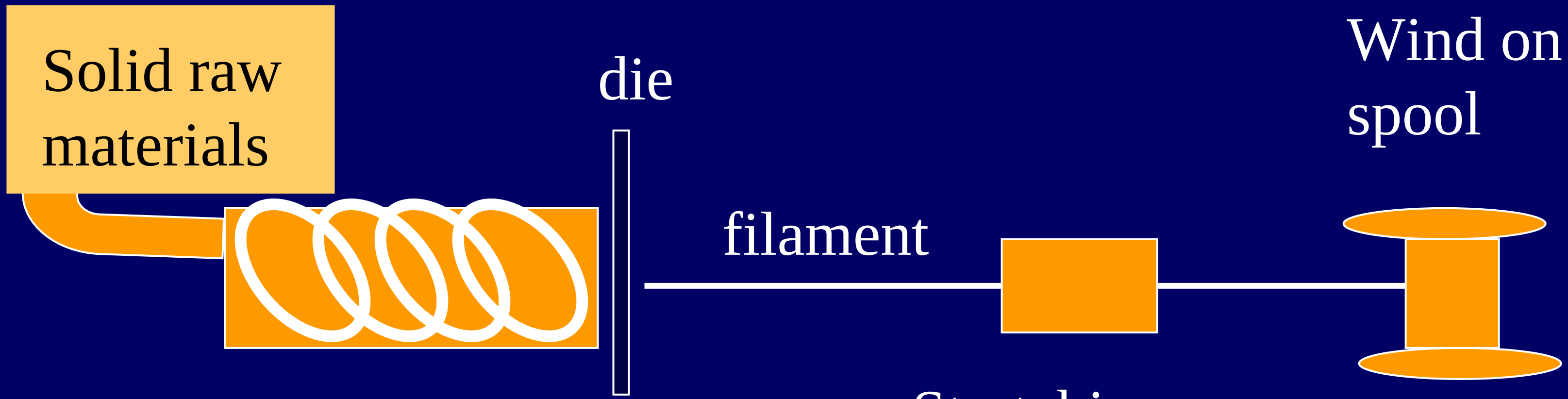
The Power of MSPC

MSPC

MSPC allows you to see relationships among variables

….and enables you to tell engineering and plant experts something new about their own system

Filament Extrusion Process



Screw Extruder: material melts from shear stresses, pressure, and heat

Molten material pushed through die

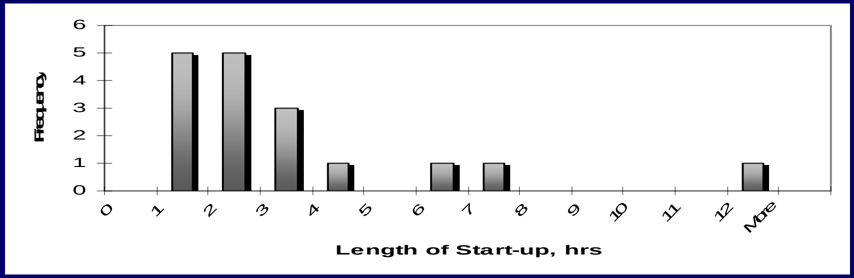
Stretching and reheating control properties like diameter and tensile strength

Sample product here – send to lab

Startup is a Problem!

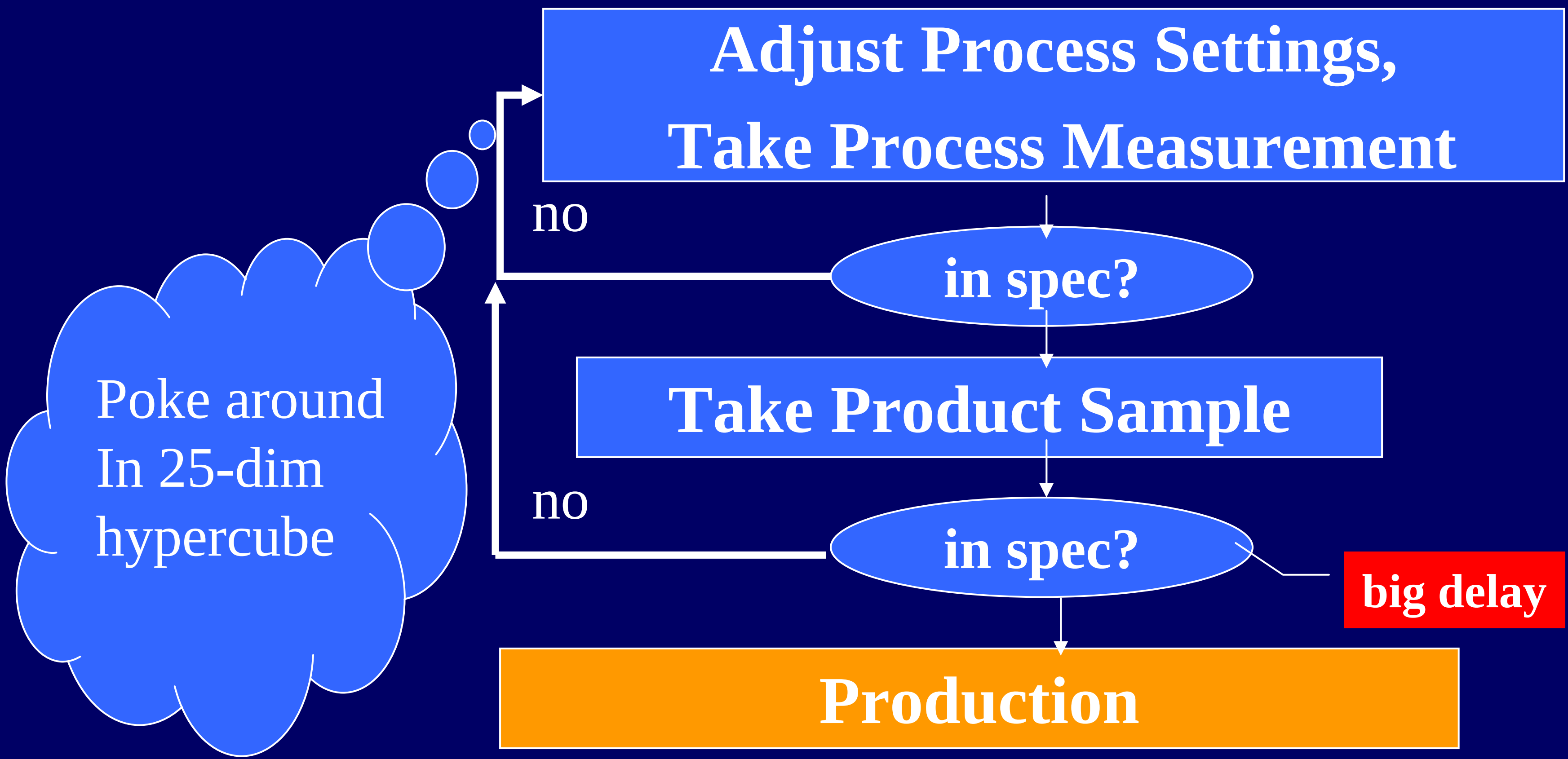
Every month: run a batch of product type

Start-up 1-12 hours
Production 12 hours



25 Process variables - data every few minutes
12 Product variables - every few hours with delay

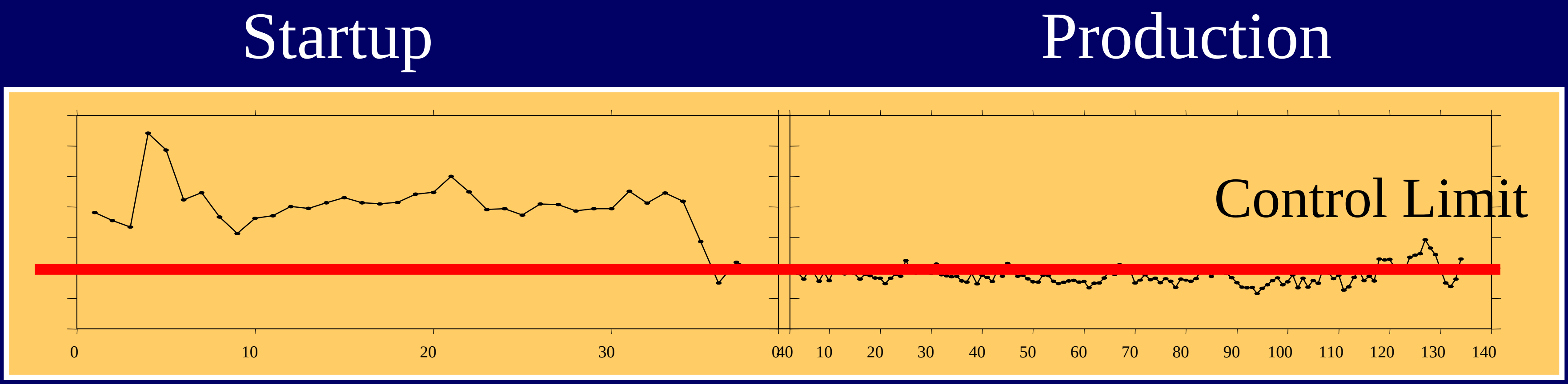
Traditional Batch Startup: Iterative & Focuses on Variables One-at-a-time



Why Different Settings for Different Batches?

- Raw materials - all in spec, but different
- Long time interval between batches
  - environment & maintenance changes

SPE for One Batch: Startup + Production

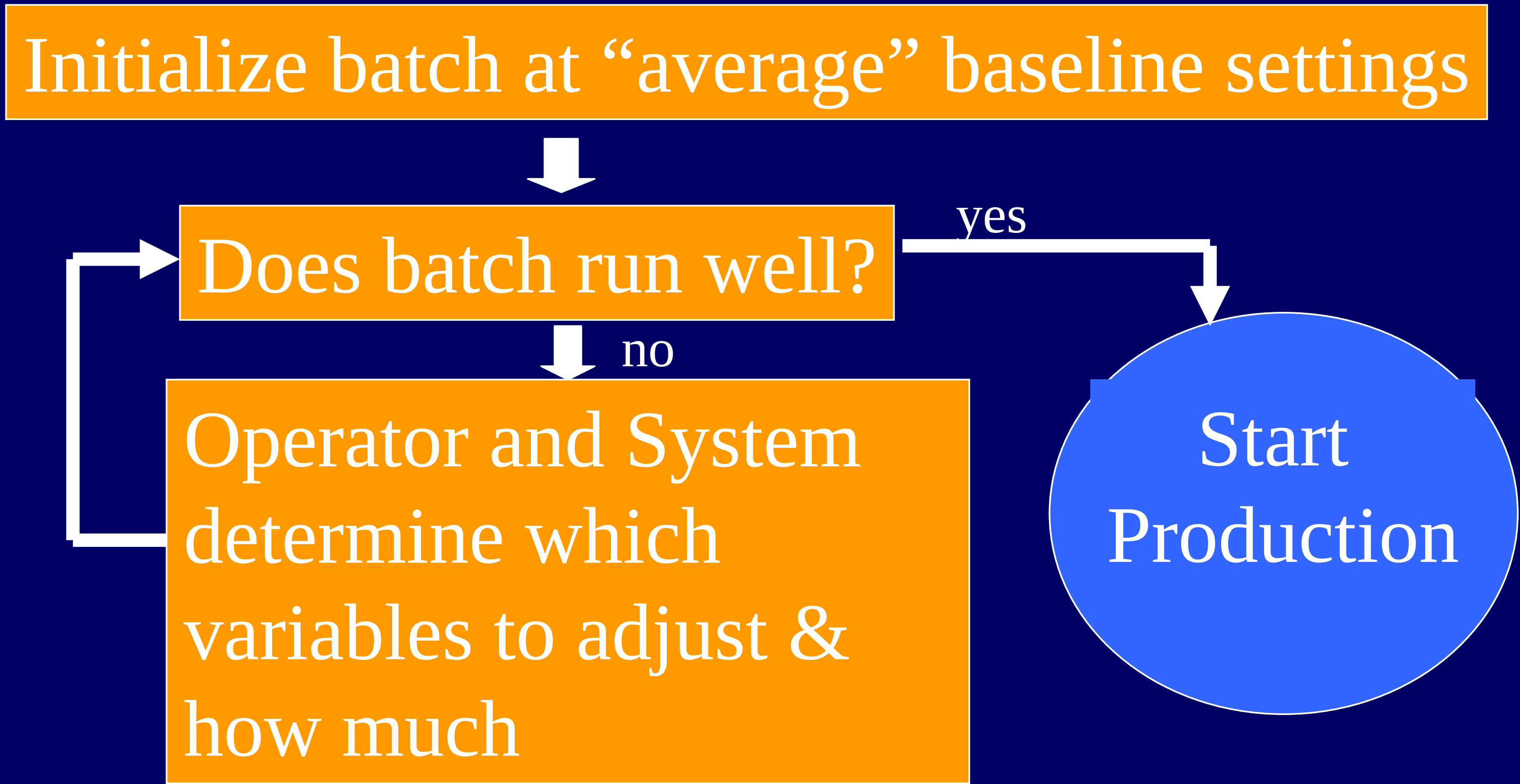


- *Initial setting inconsistent with baseline PLS model (in spec but doesn't work)*
- *No need to wait for test results of product samples*

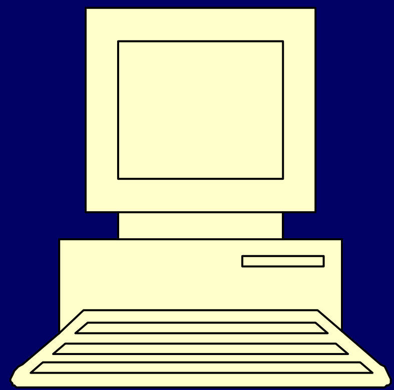
New Methodology to Reduce Start-up

- combines
  - operator input
  - optimization
  - multivariate statistics – PLS

MSPC Operator-Assisted Startup System



MSPC & Operator Work Together



Adjust v5

Set v5=xx.x
Also,
V6=yy.y
V9=zz.z

Adjust v5 or v8?

No input

Mathematical Optimization To Select New Process Settings

- Objective: minimize SPE
- Constraints:
  - Consistent with baseline model
  - Plant operation constraints
  - Not too far from current settings

Mathematical Optimization: Solve for New Process Settings

- Objective Function
  - Minimize SPE
- Subject to:
  - constraints

$$s_j^a = u$$

$$z_i = 0 \ or \ 1 \quad i = 1 \ldots k$$

$$\left| s_i^c - s_i^a \right| \le M z_i \quad = 1 \ldots k$$

$$M \ \text{large}$$

$$\sum_{i=1}^{k} z_i \le L$$

$$x^a = g(s^a)$$

$$t_i = w_i x^a \quad i = 1 \ldots A$$

$$\left| t_i \right| \le r \quad i = 1 \ldots A$$

Plant Operation Constraints – Engineering Sense

- New settings within allowable range
- Limit the size of adjustments
- Limit the number of variables adjusted
  - Introduces 0-1 variables into the mathematical optimization

Mixed Integer Quadratic Program

- Mixed decision variables
  - 0-1 variables in constraint limiting no. of adjustments
  - continuous process settings
- Objective function is convex quadratic
- Linear constraints
- Solve with Bender's Algorithm or Search

Operator Inputs Variable to Adjust & MSPC Selects Settings

- Historical
  - t=40: adjust v1 and v2
  - t=100: adjust v3
  - t=150: adjust v3 & production!

- With MSPC algorithm
  - t=40: input v2
output v1, v2, & v3
  - t=50: production!

Startup reduced 67% from 150 to 50 mins

Operator Considers 2 Adjustments & MSPC Selects

- Historical
  - –t=40: adjust v7
  - –t=60: adjust v4, v5, v6
  - –t=210: adjust v5, v6
  - –t=240: adjust v5, v6
  - –t=330: adjust v5, v6
  - –t=360: adjust v7 (start) & production
- With MSPC algorithm
  - –t=40: input v4 OR v7
output v4, v5, v6
  - –t=50: production!

Startup reduced 86% from 360 to 50 mins

MSPC Suggests Adjustments with No Prompt From Operator

- Historical
  - t=190: adjust v8
  - t=250: adjust v9 & production

- With MSPC algorithm
  - t=120: SPE is large
    output v8, v9
  - t=130: production!

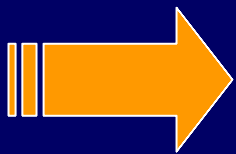
Startup reduced
48% from
250 to 130 mins

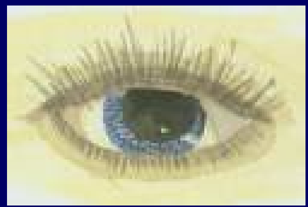
Benefits of Reduced Mean and Variance of Batch Startup Time

- reduce start-up time 40%
- improve production planning
- increase capacity
- ease bottleneck off-line testing
- reduce scrap

The Power of MSPC

MSPC





Select Optimal Process Settings
in Steel Manufacturing

Find Process Settings to Optimize Quality

- Magnetic steel sheets, 1432 observations
  - One major product variable
  - Seven major process variables
- Premium price if product quality $< C^*$
  - currently 12% of sheets
- Goal: find settings to minimize product var

Data Mining + PLS Work Together

PLS Model

---Reduces dimensionality

---Independent PLS comps

+

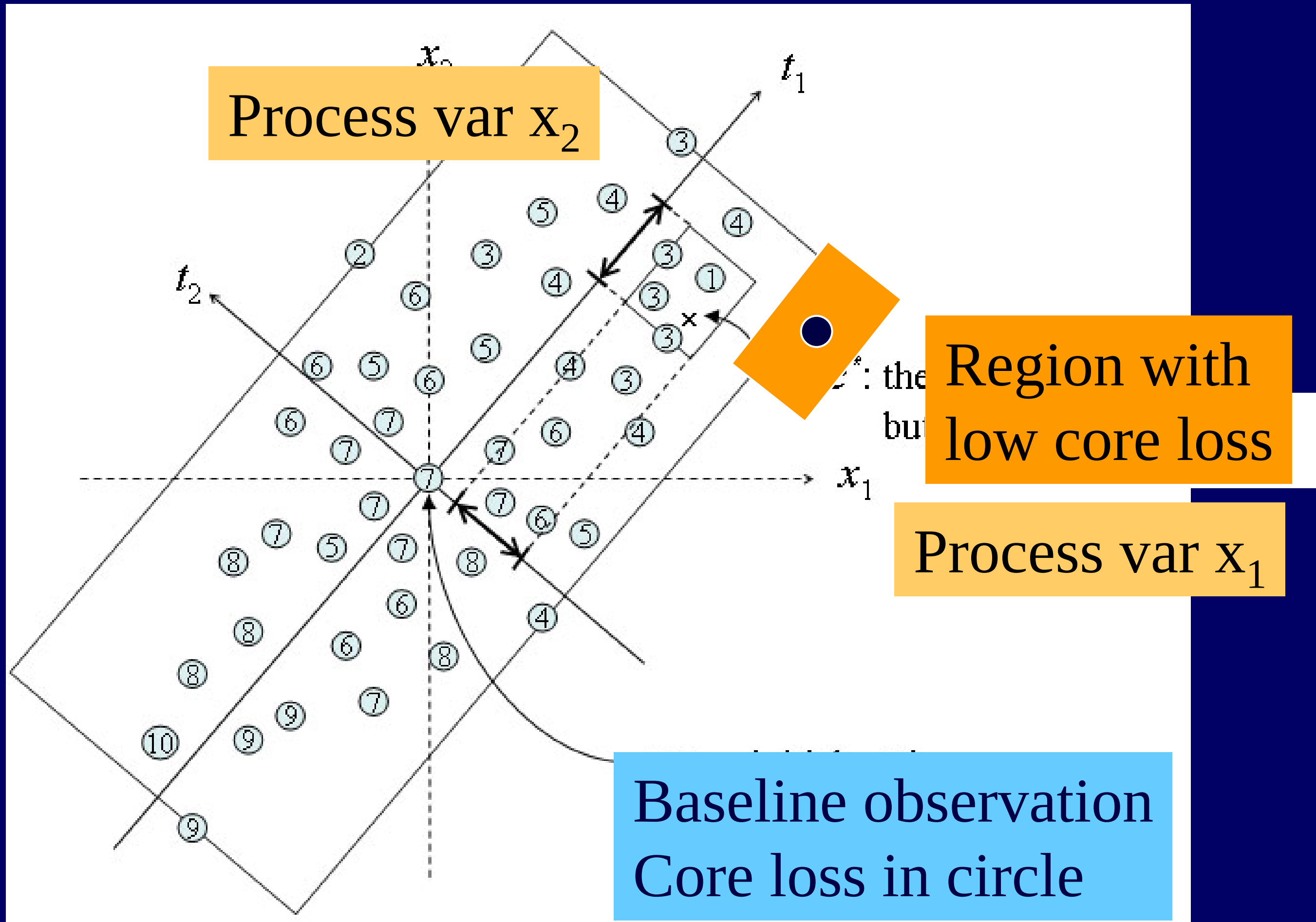
Data Mining Rule Induction Method

---Identifies best process settings

---Without constructing explicit quality function

---Must use PLS first!

Data Mining Rule Induction Method Finds Best Settings



New Settings Yield Improvement in Steel Sheet Production

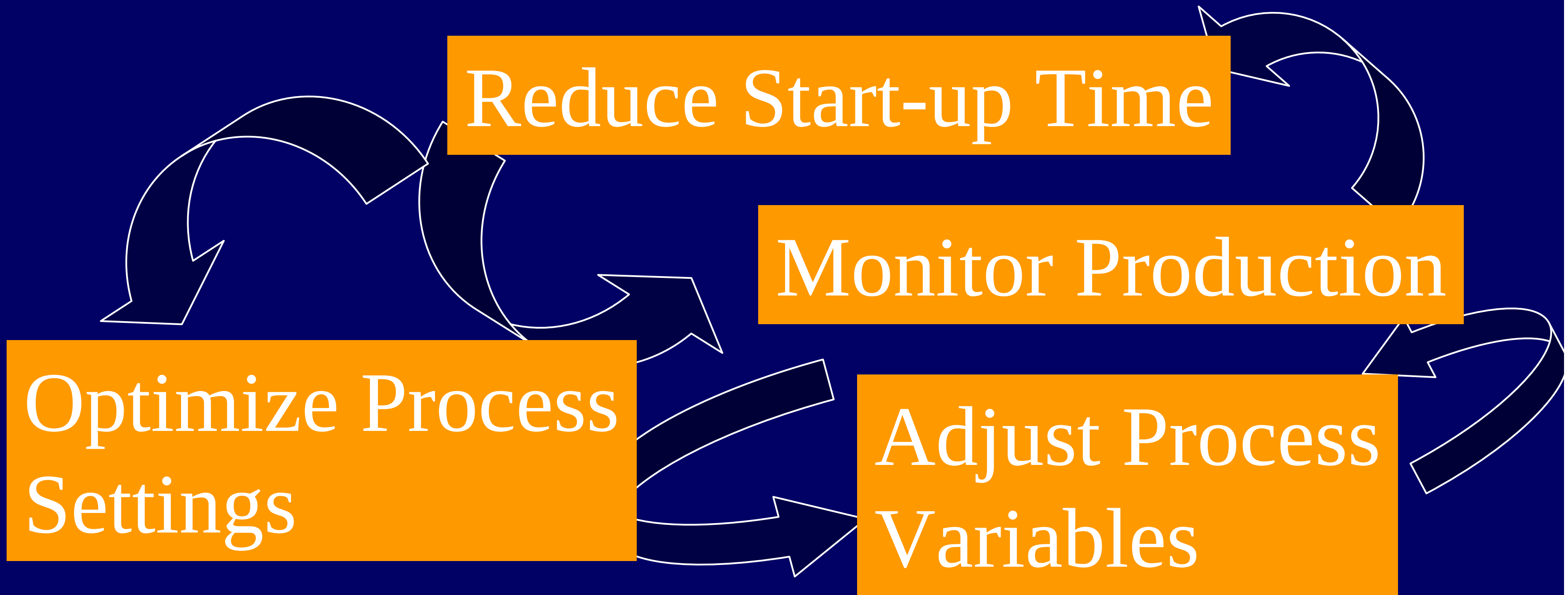
- Reduced product quality variable by 8% (lower the better)

- Increase in sheets sold at premium price:

  12% => 25%

The Power of MSPC

Model relationships among multiple, correlated process and product variables



END

Baseline Modeling: Why Use PLS Rather Than Multiple Regression?

- Use baseline to construct multiple regression model for prediction $Y=f(x)$
- Choose process settings X to maximize probability product variables Y within specs

Polystyrene Extrusion Simulation: PLS vs. Regression

- 7 process variables
- 8 productivity variables

Optimal Setting	PLS	Regression
Flow rate (kg/hr)	67.0	68.9
Screw speed (rpm)	105.3	116.3
Barrel temperature	200	212.6
% successful simulations	93%	60%

9000 Simulation Runs

PLS/ Rule Induction

99% Better quality, by 67%

1% Worse quality, 28%

Safe choice

Regress Form Known

92% Better quality, by 81%

8% Worse quality, 104%

risky – depends on coefficient estimates