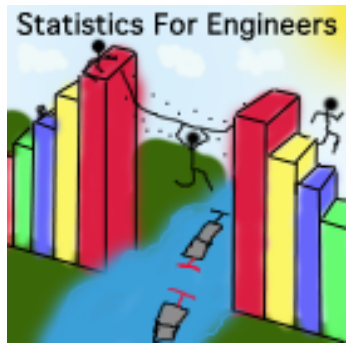
Statistics for Engineers

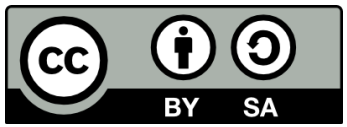
ChE 4C3 and 6C3



kevin.dunn@mcmaster.ca
http://learnche.mcmaster.ca/4C3

Overall revision number: 153 (April 2014)

Copyright, sharing, and attribution notice



This license allows you:

- **to share** - to copy, distribute and transmit the work, including print it
- **to adapt** - but you must distribute the new result under the same or similar license to this one
- **commercialize** - you *are allowed* to use this work for commercial purposes
- **attribution** - but you must attribute the work as follows:
  - "Portions of this work are the copyright of Kevin Dunn", *or*
  - "This work is the copyright of Kevin Dunn"

  (when used without modification)

We appreciate:

- if you let us know about **any errors** in the slides
- **any suggestions to improve the notes**

All of the above can be done by writing to

`kevin.dunn@mcmaster.ca`

or anonymous messages can be sent to Kevin Dunn at

http://learnche.mcmaster.ca/feedback-questions

If reporting errors/updates, please quote the current revision number: 153

Please note that all material is provided "as-is" and no liability will be accepted for your usage of the material.

In context

Sections covered so far:

1. Visualizing data
2. Univariate statistics
3. Combine those two areas: create a system to visually monitor any process

AIM: **rapid problem detection**

- then comes diagnosis (cover this later)
- and process adjustment/fixing (not covered)

Note: this is a tool to assist *Troubleshooting*

Examples

Systems you may have seen:

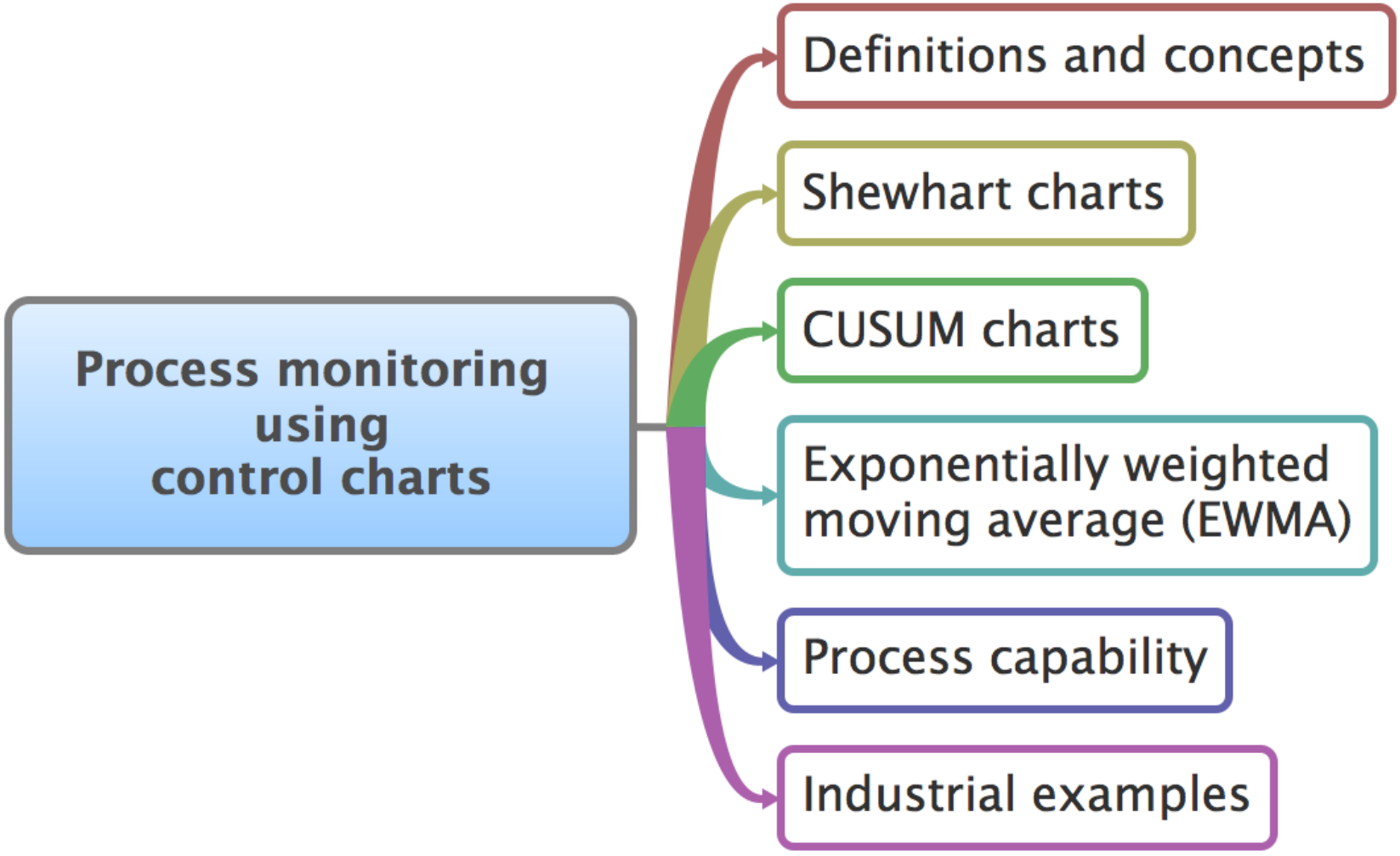
- hospital (monitoring patients)
- stock market charts (intraday trading)
- processing/manufacturing facility
- You will see soon: The ExCEL building

Examples:

- *Co-worker*: are our product dimensions (or some other production quality measurement) stable?
- *Yourself*: how can we quickly detect a slow drift in the process?
- *Manager*: track the hourly average profit, and process throughput and react to any problems.
- *Engineer*: we can show operators the data in an efficient way, so they can move the process away from unsafe operation
- *Potential customer*: what is your process capability? We are looking for Cpk of at least 1.6.

Note: process monitoring is mostly **reactive** and not *proactive*. So it is suited to *incremental* process improvement.

What we will cover and some references for you to consider



Concepts: ARL, LCL, UCL, Cpk, Shewhart charts, EWMA, CUSUM, Type I and II errors, false alarms, real-time applications

What we will cover and some references for you to consider

The standard "process control" textbooks have chapters on basics of Process Monitoring:

- Marlin, "Process Control", 2nd edition, Chapter 26
- Seborg, Edgar, Mellichamp and Doyle: "Process Dynamics and Control", 3rd edition, Chapter 21

What is process monitoring about?

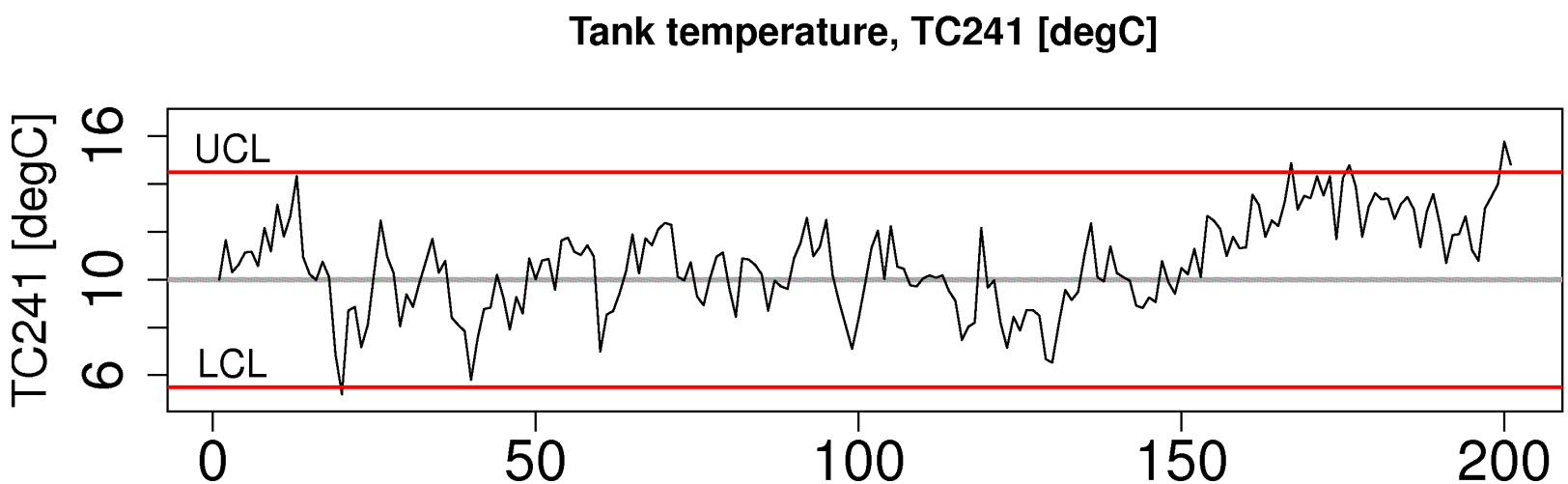
- We know that quality is not optional; customers move onto suppliers that provide quality products
- Quality is generally not a cost-benefit trade-off
  - Customer's value quality; but it is a long-term benefit
  - Example: car sales in North America: the steady rise of the Asian manufacturers; starting to change ...

Control charts

Used to display and detect this unusual variability

- it is most often a time-series plot, or sequence plot
- a target value should be shown
- one or more limit lines are shown
- displayed in real-time, or pretty close to real-time
- usually applied on many important process variables
- displayed in the units of the variable

Example:



Control charts: demonstration

Process monitoring: relationship to feedback control

- Also called "Statistical Process Control" (SPC)
- We will avoid this term due to potential confusion
- However you will see the term "control chart" = monitoring chart
- Monitoring is *similar* to (feedback) control:
  - it is continually applied
  - we check for deviations (error)
- Monitoring is *different* to (feedback) control:
  - no action taken, unless required
  - i.e. adjustments are **infrequent**
  - action taken is usually **manual**
  - adjust due to **special causes**
- Process monitoring: make *permanent* adjustments to reduce variability
- Feedback control: *temporarily* compensates for the problem all the time

General approach

- **Phase 0**: decide what to monitor
- **Phase 1**: building and testing from off-line data
  - very iterative: remove outliers
  - calculate limits, test if they are useful, repeat
  - you will spend most of your time here
- **Phase 2**: using the control chart
  - on new, unseen data
  - implemented with computer hardware and software
  - usually for real-time display

Phase 0: Decide what to monitor?

Discuss these in groups: what would you monitor

- Waste water treatment process
- Tablet/pharmaceutical manufacturing
- Oil and gas (e.g. a distillation column)
- Food-processing unit (e.g. a fryer)
- Mineral processing plant (e.g. a flotation cell)
- Plastics processing (e.g. a twin-screw extruder)

In-control vs out-of-control

"In-control": behaviour of the process is stable over time. Also called *common cause* operation.

The opposite is: out of control, assignable causes, special causes, destabilizing event, off-target.

Shewhart chart

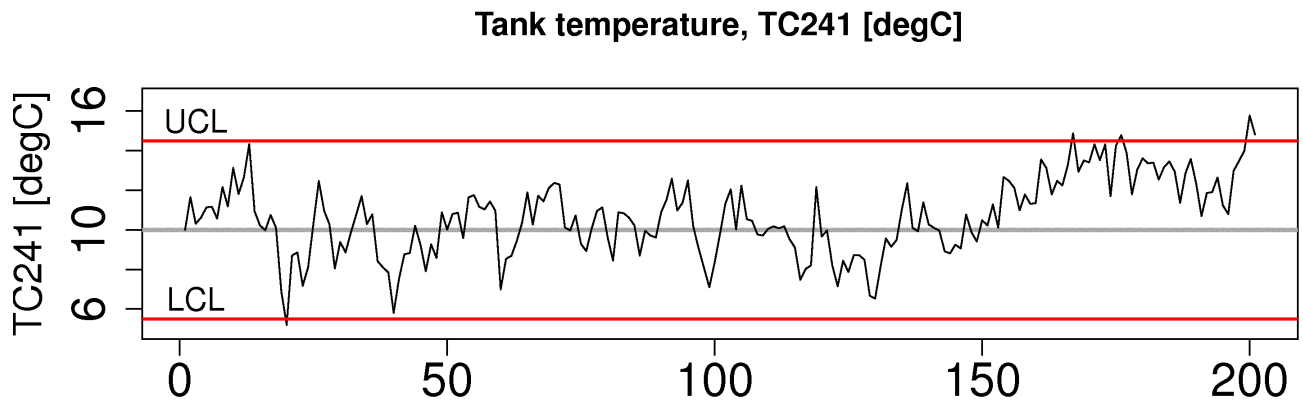
- Named for *Walter Shewhart* from Bell Telephone and Western Electric, parts manufacturing, 1920's
- A chart for monitoring variable's *location*

It has:

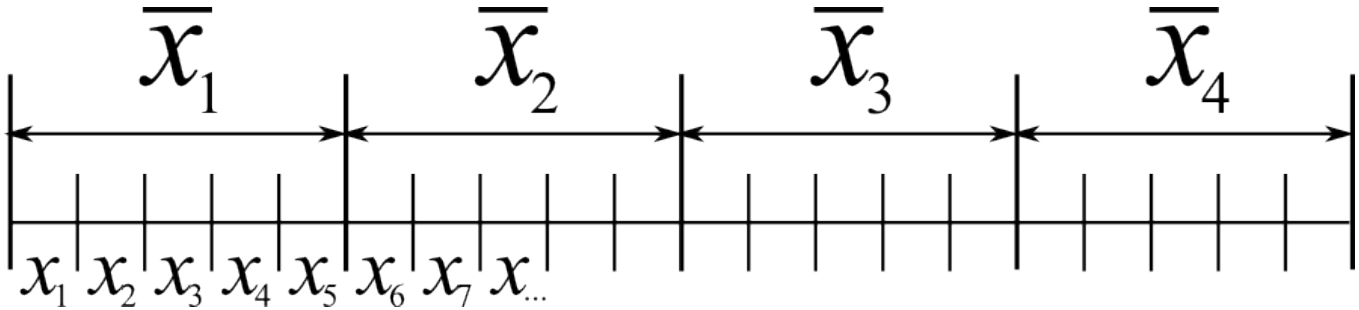
- lower control limit (LCL)
- upper control limit (UCL)
- a target

No action taken if the variable plotted remains within limits.

Do not tamper with the process if you are in control!



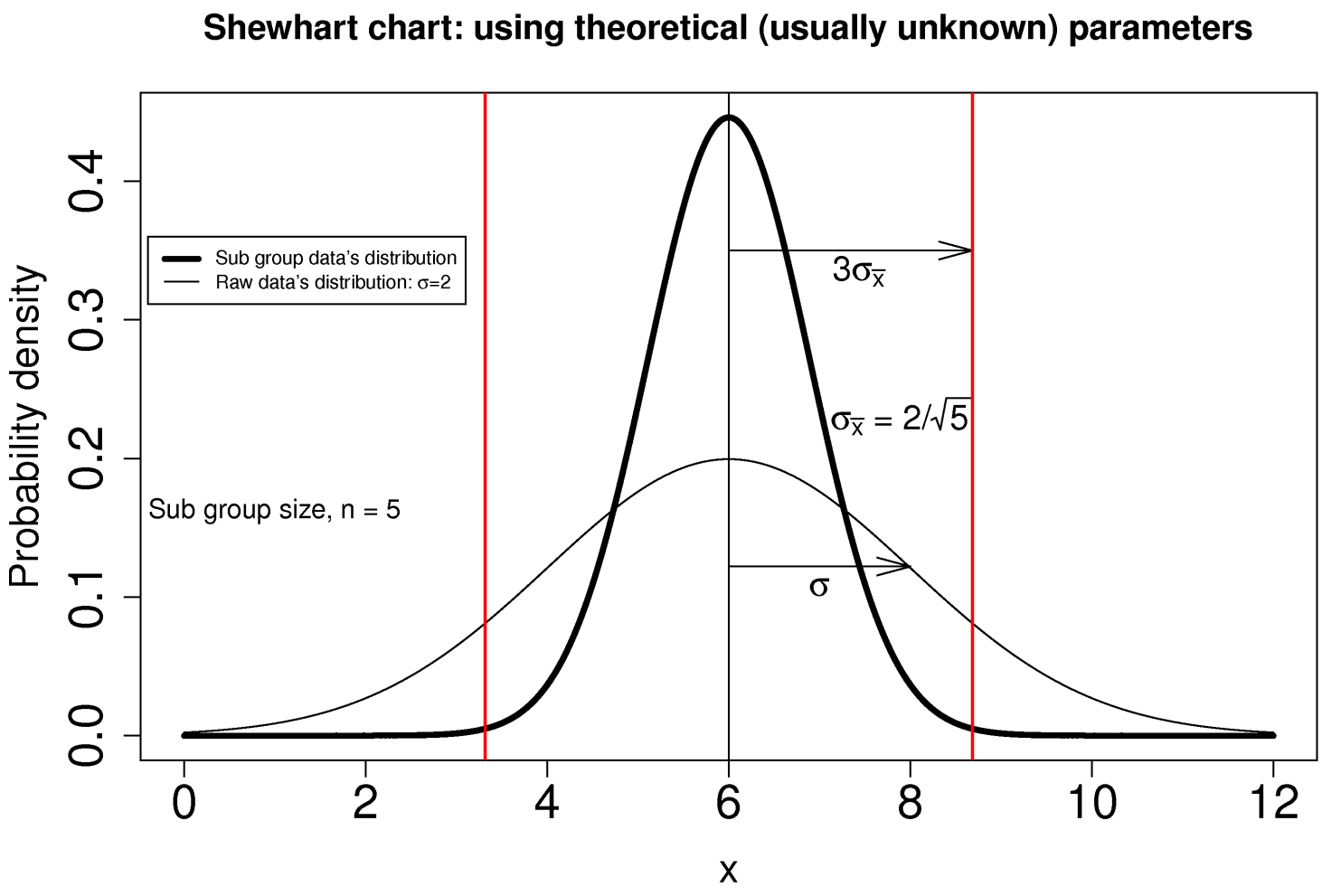
Shewhart chart: Derivation for the limits (LCL and UCL)



- Take a subgroup of samples, of size n ($n = 5$ in the picture here)
- You are free to select the value of n, the subgroup size
- Calculate $\bar{x}$ from the n values
- What is the distribution of $\bar{x}$?
  - $\bar{x} \sim \mathcal{N}(\mu, \sigma^2/n)$
- Define: $\sigma_{\bar{X}} = \sigma/\sqrt{n} =$ std deviation of my subgroup average
- Assume we know μ (that's the target line) and σ
- σ is used to find the LCL and UCL

Shewhart chart: Derivation for the limits (LCL and UCL)

- **Thin line**: single process measurements have $\mu = 6$ and $\sigma = 2$
- **Thick line**: $\sigma_{\overline{x}} = 2/\sqrt{5} = 0.894$, because $n = 5$
  - Upper bound $= \mu + 3\sigma_{\overline{x}} = 6 + 3 \times 0.894 = 8.68$
  - Lower bound $= \mu - 3\sigma_{\overline{x}} = 6 - 3 \times 0.894 = 3.318$



Shewhart chart: Derivation (theoretical process)

- z-value: $z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{X}}}$
- Confidence interval for μ: $\bar{x} - c_n\sigma_{\bar{X}} < \mu < \bar{x} + c_n\sigma_{\bar{X}}$
- By convention we use $c_n = 3.0$
- Area between LCL and UCL: 99.73%
- A chance of 1 in 370 that a data point, $\bar{x}$, will lie outside these bounds
- The bounds are for $\bar{x}$, not for individual, raw x values

Critical limitation

All of the above assumed we knew μ and σ

Shewhart chart: using estimates

But, we don't know population parameters

- For μ: use $\overline{\overline{x}} = \frac{1}{K}\sum_{k=1}^{K}\overline{x}_k$; where K = number of phase 1 groups
- For μ: or just use known target value
- For μ: or use the **median** of a long sequence of data
- For σ:
  - Define: s_k = standard deviation of n values
  - Define: $\overline{S} = \frac{1}{K}\sum_{k=1}^{K} s_k$
  - Then overall standard deviation is estimated from $\hat{\sigma} = \frac{\overline{S}}{a_n}$, where a_n is a correction factor

n	2	3	4	5	6	7	8
a_n	0.798	0.886	0.921	0.940	0.952	0.959	0.965

$$\text{limits} = \overline{\overline{x}} \pm 3 \cdot \frac{\sigma}{\sqrt{n}} \qquad \text{LCL} = \overline{\overline{x}} - 3 \cdot \frac{\overline{S}}{a_n\sqrt{n}} \qquad \text{UCL} = \overline{\overline{x}} + 3 \cdot \frac{\overline{S}}{a_n\sqrt{n}}$$

Example

We measure 5 colour values on each rubber bale:

- e.g. bale 1 raw values: [231, 251, 235, 241, 227]
  - $\bar{x}_1 = 237$ is our data point for the Shewhart chart
  - $s_1 = 9.38$ is the std dev for the 5 points
- e.g. bale 2 raw values: [252, 253, 247, 232, 244]
  - $\bar{x}_2 = 245.6$ and $s_2 = 8.44$

Data from 20 such $\bar{x}_k$ calculations (the 20×5 raw values are not shown):

```
[245, 239, 239, 241, 241, 241, 238, 238, 236, 248,
 233, 236, 246, 253, 227, 231, 237, 228, 239, 240]
```

Calculated for you: $\bar{\bar{x}} = 238.8$ and $\bar{S} = 9.28$

Phase 1 workflow: *build* the monitoring chart

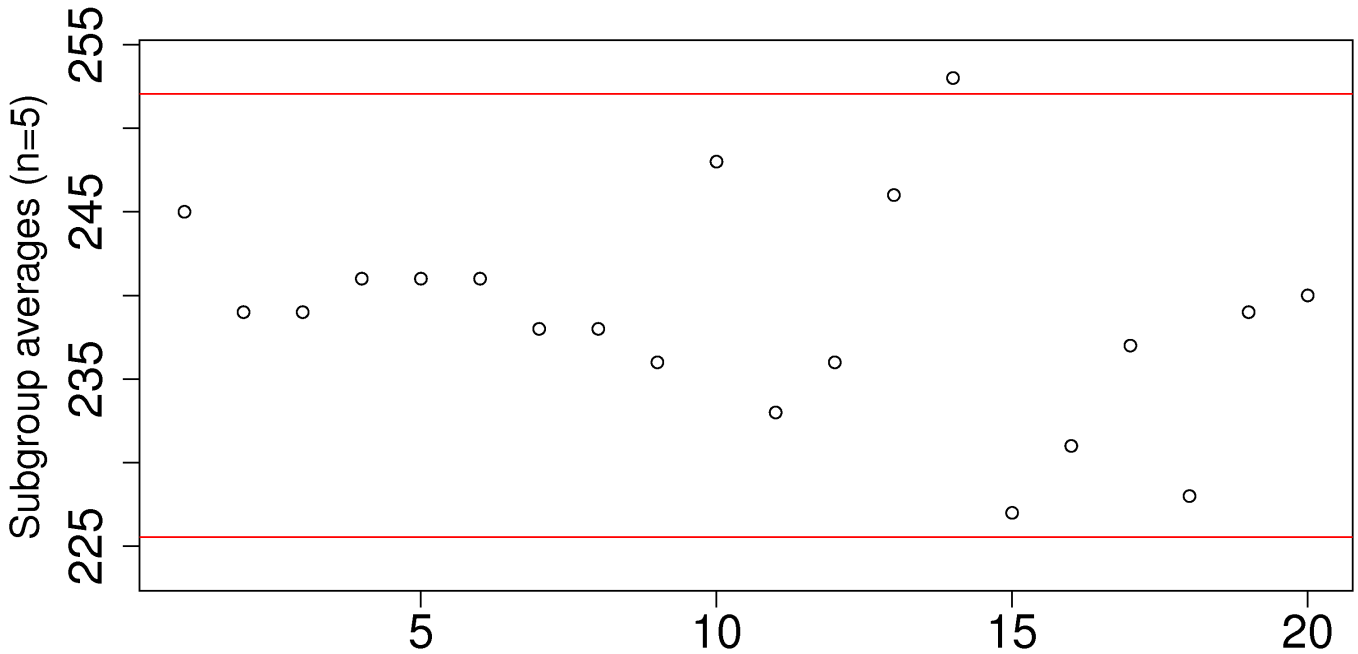
- Calculate: LCL and UCL
- Any $\bar{x}$ points outside limits?
- If so, exclude these outliers and recalculate limits.

Phase 2 workflow: *using* the monitoring chart

- Obtain n new data points
- Calculate the subgroup average, $\bar{x}$
- Add this $\bar{x}$ point to the plot

Example

- $\text{LCL} = 238.8 - 3 \cdot \dfrac{9.28}{(0.94)(\sqrt{5})} = 225.6$
- $\text{UCL} = 238.8 + 3 \cdot \dfrac{9.28}{(0.94)(\sqrt{5})} = 252.0$
- Sample with value of $\bar{x}_i = 253$ exceeds these limits



- After excluding it:
  - new $\bar{\bar{x}} = 238.0$, and new $\bar{S} = 9.68$
  - new LCL $= 224$, new UCL $= 252$

Back to the bigger picture

We have covered this:

- **Phase 0**: decide what to monitor
- **Phase 1**: building and testing from off-line data
  - very iterative: remove outliers
  - calculate limits, test if they are useful, repeat
  - you will spend most of your time here

We are here now:

- **Phase 2**: using the control chart
  - on new, unseen data
  - implemented with computer hardware and software
  - usually for real-time display

Assessing the chart's performance: error probability

- **Type I error**: $\bar{x}$ typical of normal operation, but falls outside UCL or LCL limits
  - Theoretical derivation: that happens 1 in 370 times (99.73% inside limits)
  - Also called $\alpha = 0.0027$ when using $\pm 3\sigma_{\bar{x}}$ limits ($0 < \alpha < 1$)
  - *Synonyms*: false alarm, false positive (diseases), producer's risk (acceptance sampling), false reject rate
- **Type II error**: $\bar{x}$ is not stable, but falls within UCL and LCL limits
  - Called β: is a function of the degree of difference (next)
  - *Synonyms*: false negative, consumer's risk, false acceptance rate
- Asymmetrical risks: airport screening, disease diagnosis, trial by jury, making drugs
- In pseudo-math:

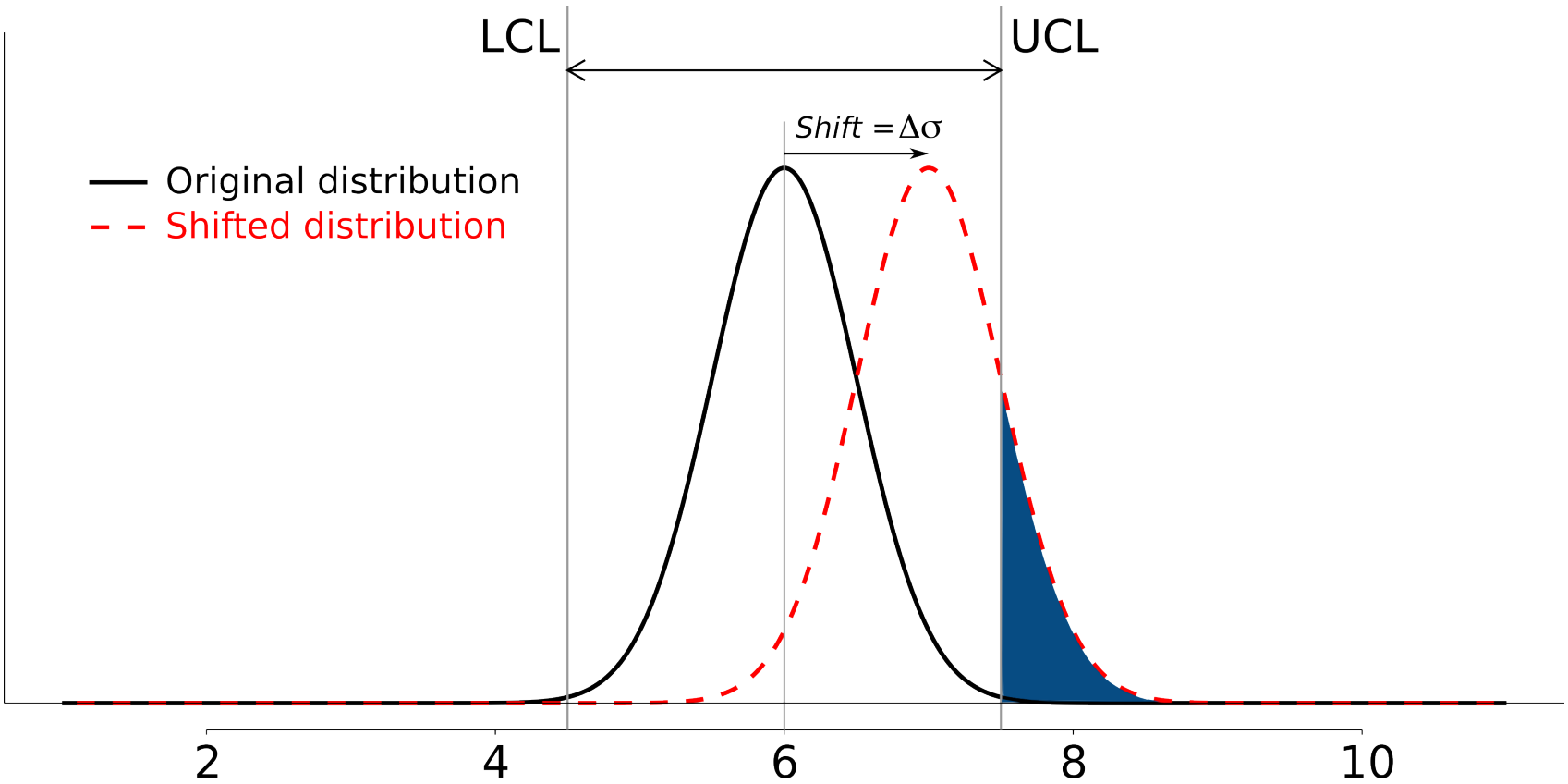
$$\alpha = Pr\,(\overline{x} \text{ is in control, but lies outside the limits})$$
$$\beta = Pr\,(\overline{x} \text{ is not in control, but lies inside the limits})$$

Assessing the chart's performance: type II error

Assume $\bar{x}$ from shifted distribution: μ to $\mu + \Delta\sigma$

Question: what is the probability a new $\bar{x}$ will fall within existing LCL and UCL?

- *Answer*: 1 − shaded area



Assessing the chart's performance: type II error

Δ	0.25	0.50	0.75	1.00	1.50	2.00
β when $n = 4$	0.9936	0.9772	0.9332	0.8413	0.5000	0.1587

- β is a function of the process shift, Δ
- Table is for $n = 4$ and UCL and LCL at the $\pm 3\sigma_{\bar{X}}$ limits
- `beta <- pnorm(3 - delta*sqrt(n)) - pnorm(-3 - delta*sqrt(n))`
- Interpretation: Shewhart chart not good at detecting small to medium changes in the location!
- Surprising, given that we use Shewhart chart for this purpose.
- E.g: a 0.75σ only be detected around (1 - 0.9332) = 6.7% of the time

Adjusting the chart's performance

Key point

Control chart limits are not set in stone. Adjust them!

Nothing makes a control chart more useless to operators than frequent false alarms.

- α: simply move LCL and UCL up and down (not incorrect to do this!)
- β: as you increase UCL, $\alpha \to 0$, but $\beta \to 1$
- But note that as you decrease type I error, your type II error will increase
- Cannot simultaneously have low type I and type II error

Reducing the control limits

After making permanent changes to the process to eliminate the problem:

- you might notice the process variability has decreased (*how will you pick that up*?)
- if you can maintain that reduced variability, then **tighten** the control limits
- to maintain the gains you have made

Average run length (ARL)

- ARL = average number of sequential samples we expect before seeing a point outside limits
- ARL $= \frac{1}{\alpha}$
- ARL for in control process, with 3-sigma limits?
  - $\alpha = 0.0027$ (0.27% of false alarms)
  - ARL $= 1/0.0027 = 370$

Extensions: Western Electric Rules

- Basic Shewhart chart is not too sensitive to process shifts. So supplement control limits with these additional heuristic "rules". Raise an alarm when:
  - 2 out of 3 points lie beyond 2σ on the same side of the target
  - 4 out of 5 points lie beyond 1σ on the same side of the target
  - 8 successive points lie on the same side of the center line

The theoretical ARL is reduced by using these rules

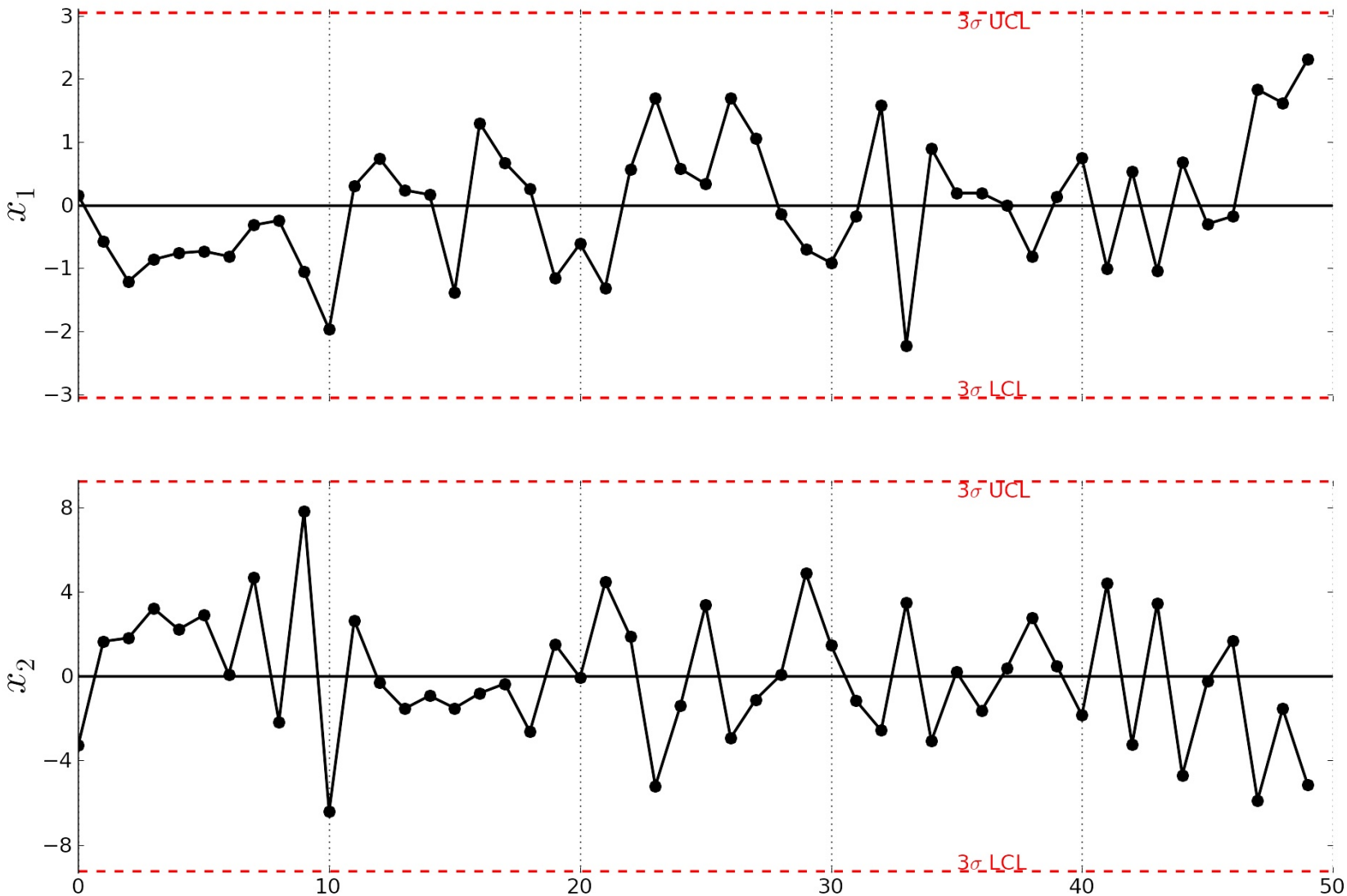
- **Adding robustness**: use robust methods in phase 1: see notes for a journal reference.

- **Warning limits**:
  - warning at $\pm 2\sigma$ (orange coloured lines or background)
  - action at $\pm 3\sigma$ (red coloured lines or background)

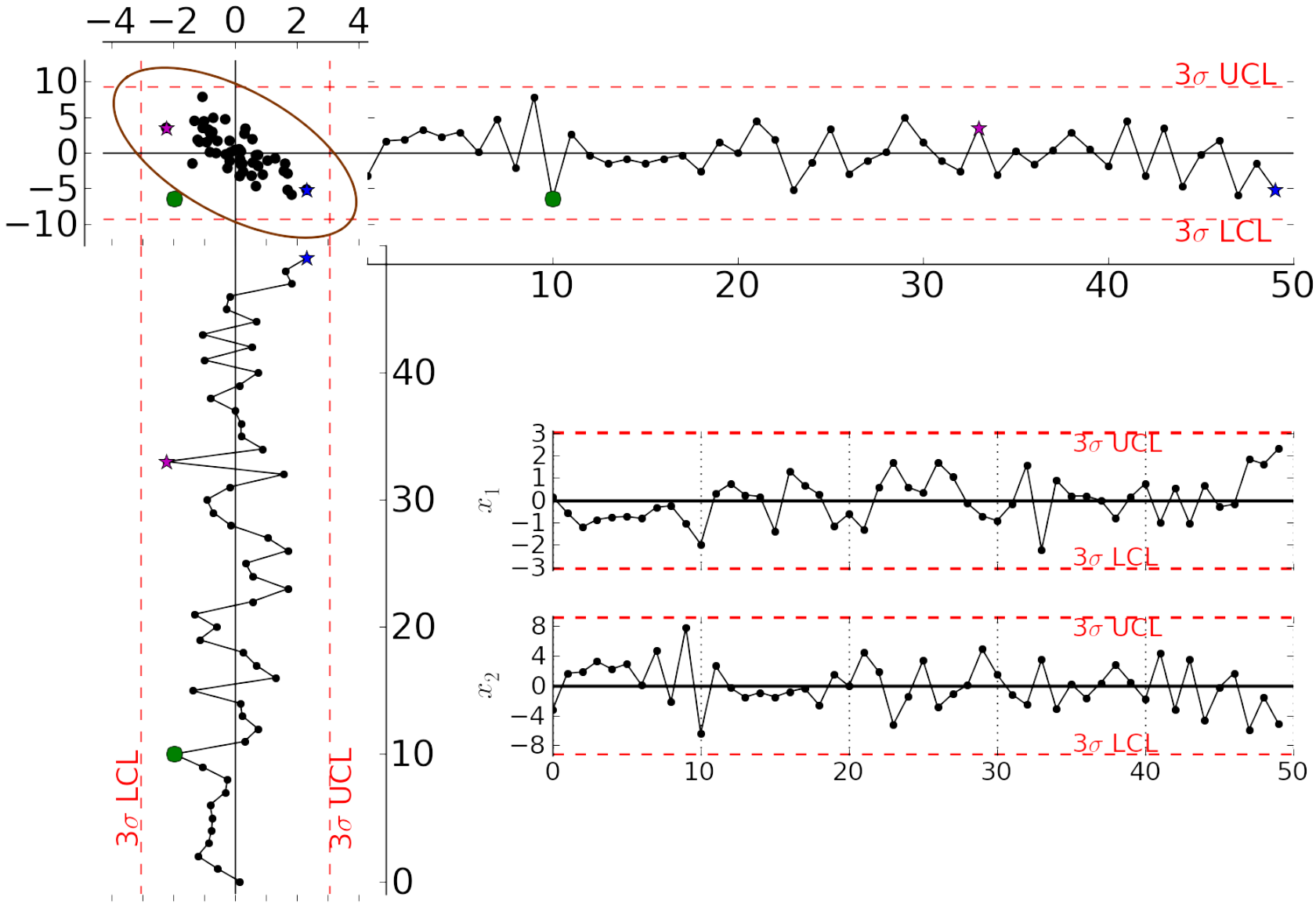
Mistakes to avoid

1. Adding product specification to the monitoring chart:
   - don't use spec limits instead of LCL and UCL
   - we are monitoring for stability, not defects (out-of-spec product)
2. Variables with heavy autocorrelation: you get much longer ARL when a fault occurs
   - use EWMA chart (next section)
3. Shewhart charts on **highly correlated quality variables**
   - an important topic - introduced in latent variable section
   - also see next two slides for a quick explanation

What is the relationship between these variables?



Monitoring correlated variables: we need a better tool



CUSUM (**cu**mulative **sum**) charts

- Shewhart chart takes a long time to detect shift in the mean, away from target, T
- CUSUM formula:

$$\begin{aligned} S_0 &= (x_0 - T) \\ S_1 &= (x_0 - T) + (x_1 - T) = S_0 + (x_1 - T) \\ S_2 &= (x_0 - T) + (x_1 - T) + (x_2 - T) = S_1 + (x_2 - T) \end{aligned}$$

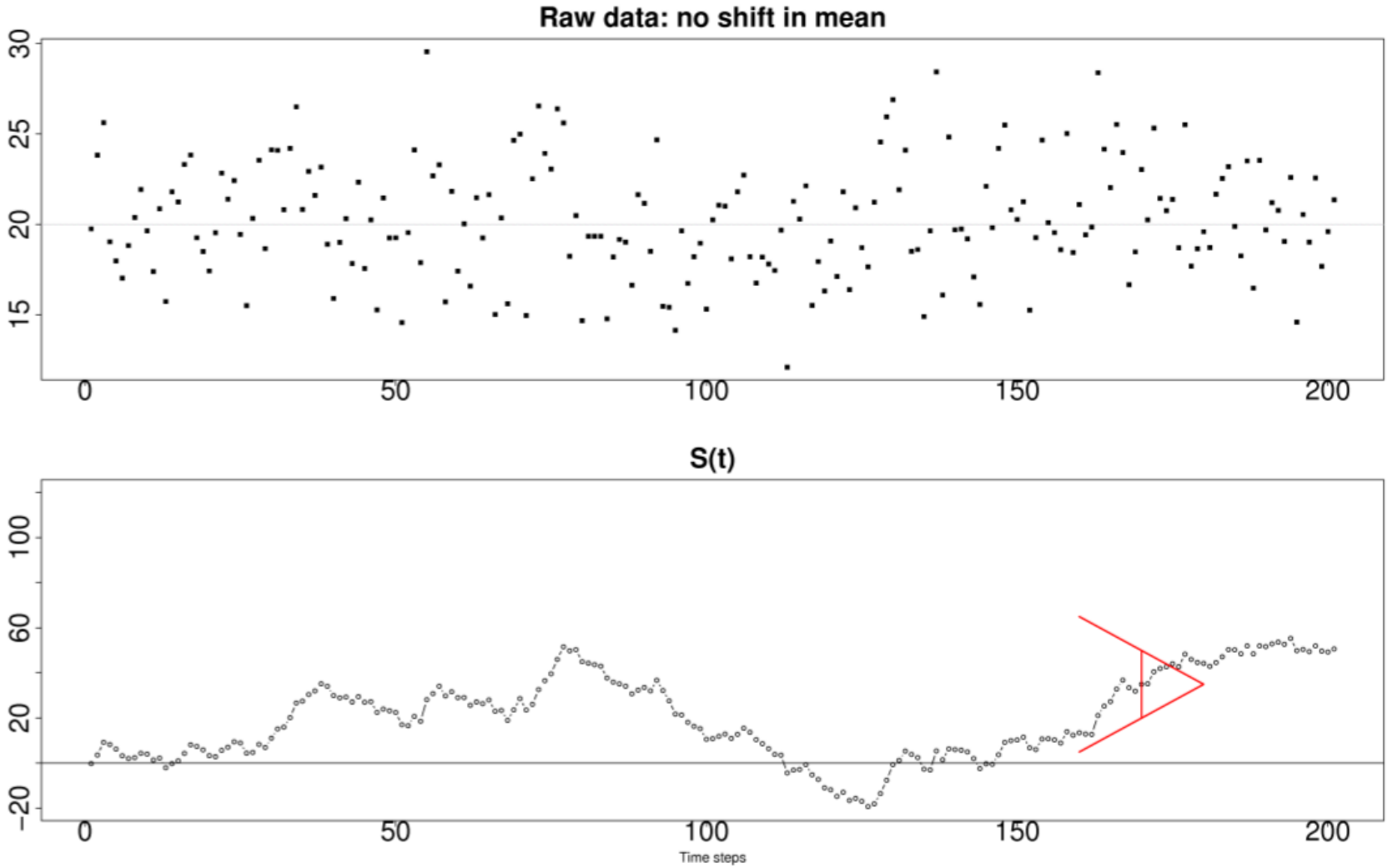
CUSUM formula:

$$S_t = S_{t-1} + (x_t - T)$$

- For process shifts: we are adding Δ to every x_t
- Accumulates: creates a steep up or down slope

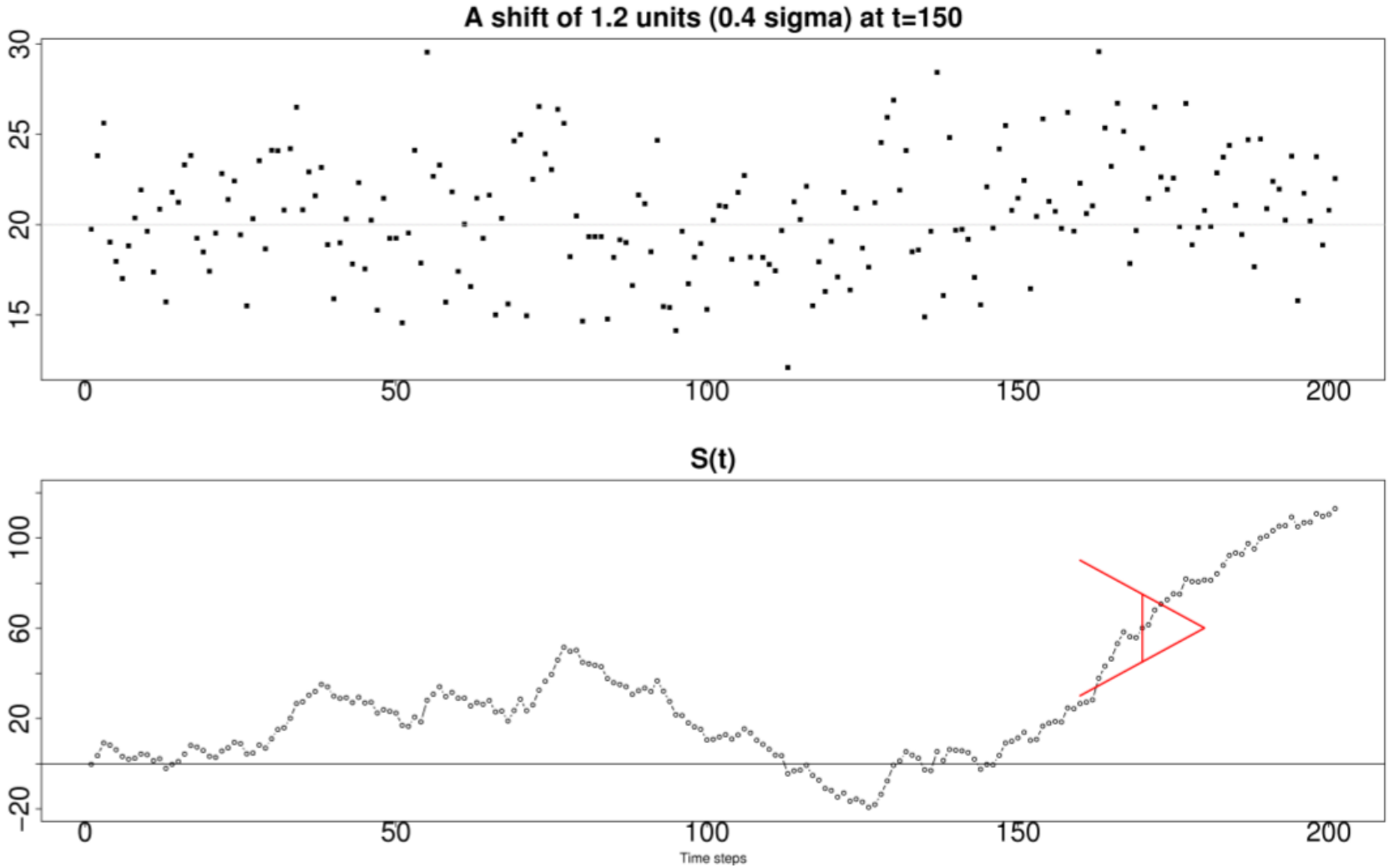
CUSUM charts: everything is OK, "in-control"

- $\mu = 20$ and $\sigma = 3$



CUSUM charts: "out-of-control"

- Shift of 1.2 units starts at $t = 150$, caught at around $t = 180$

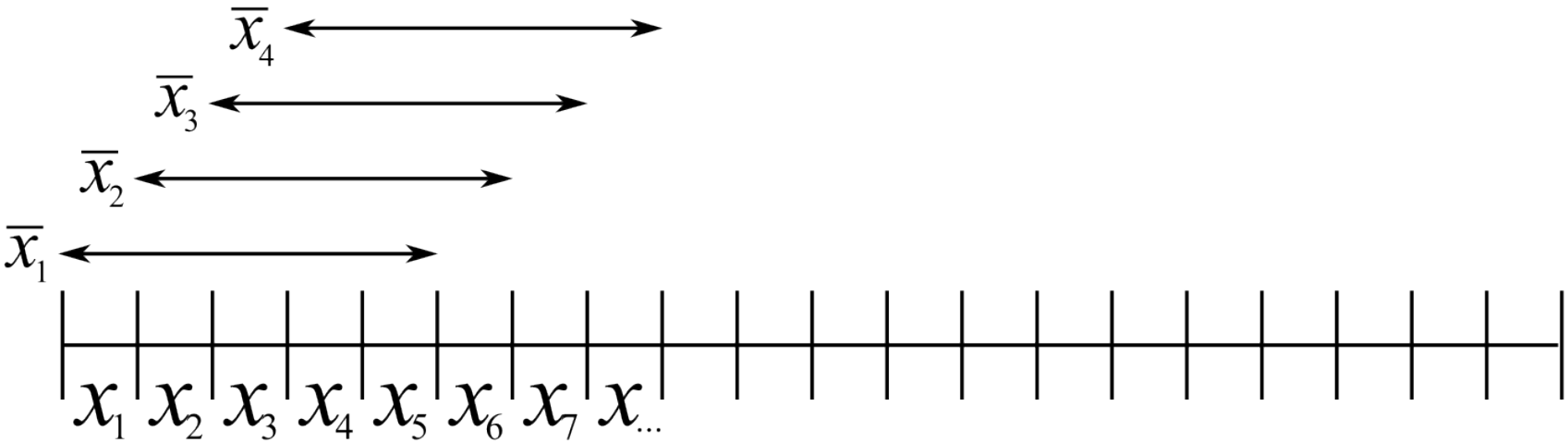


Using a CUSUM chart

- Type I and II error set by the angle and distance of the V-mask
- Implemented by computers
- If a fault is detected, reset S_t to a new value and restart chart

EWMA charts: what is a moving average?

1. Shewhart: each subgroup is independent (unrelated), no "memory"
2. CUSUM: infinite build up of errors, all the way back to $t = 0$
3. Moving average (MA) chart: has a "window" of memory:



$$\bar{x}_t = \frac{1}{n}x_{t-1} + \frac{1}{n}x_{t-2} + \ldots + \frac{1}{n}x_{t-n}$$

$$\bar{x}_t = 0.25x_{t-1} + 0.25x_{t-2} + 0.25x_{t-3} + 0.25x_{t-4} + 0 \qquad \text{for n=4}$$

- essentially gives equal "weight" of $\frac{1}{n}$ to every raw data point

EWMA derivation

- Exponentially Weighted Moving Average = EWMA
  - don't give equal weights, rather ...
  - heavier weights for the most recent observations
  - small weights further back in time

$$
\begin{aligned}
x_t &= \text{new data at time step } t \\
\hat{x}_t &= \hat{x}_{t-1} + \lambda e_{t-1} \\
e_t &= x_t - \hat{x}_t
\end{aligned}
$$

$$
\hat{x}_{t+1} = \hat{x}_t + \lambda e_t
$$

To start if off:

- $\hat{x}_0 = T$
- $e_0 = 0$
- $\hat{x}_1 = T$

And: $0 \leq \lambda \leq 1$

EWMA derivation

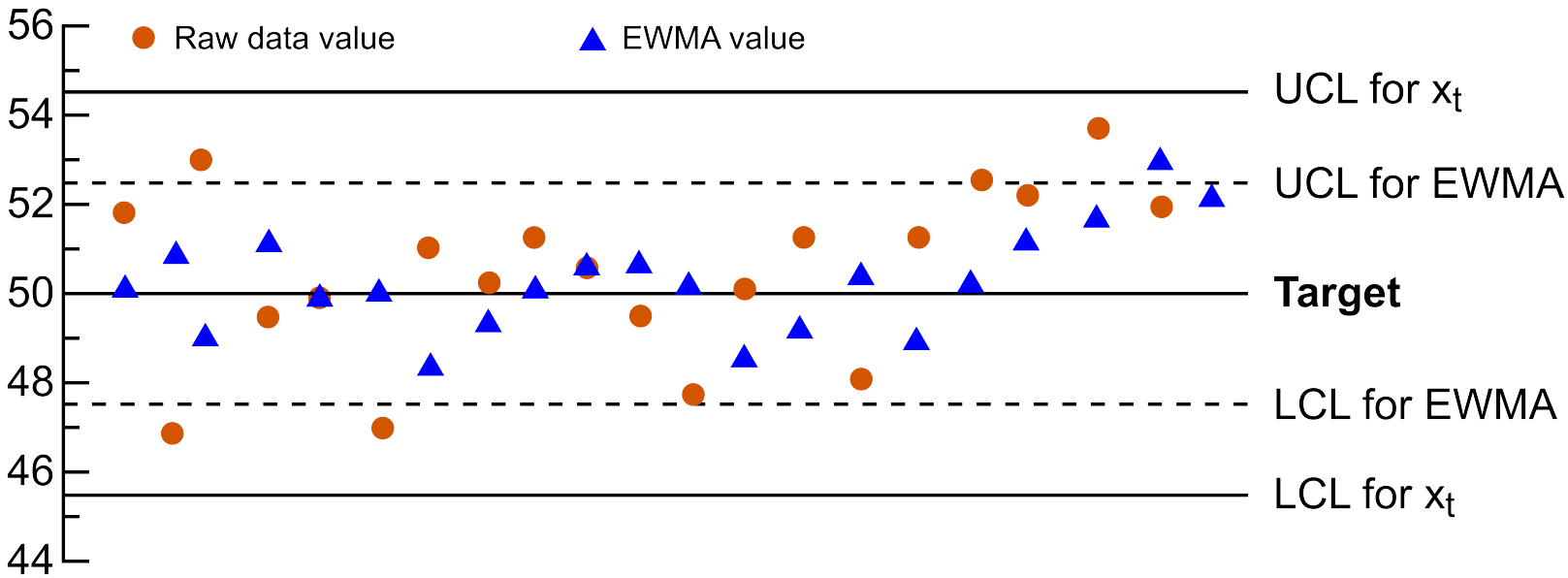
$$
\begin{aligned}
x_t &= \text{new data at time step } t \\
\hat{x}_{t+1} &= \hat{x}_t + \lambda e_t \\
e_t &= x_t - \hat{x}_t
\end{aligned}
$$

Substitute in the e_t :

$$
\begin{aligned}
\hat{x}_{t+1} &= \hat{x}_t + \lambda \left(x_t - \hat{x}_t \right) \\
\hat{x}_{t+1} &= \lambda x_t + \left(1 - \lambda \right) \hat{x}_t
\end{aligned}
$$

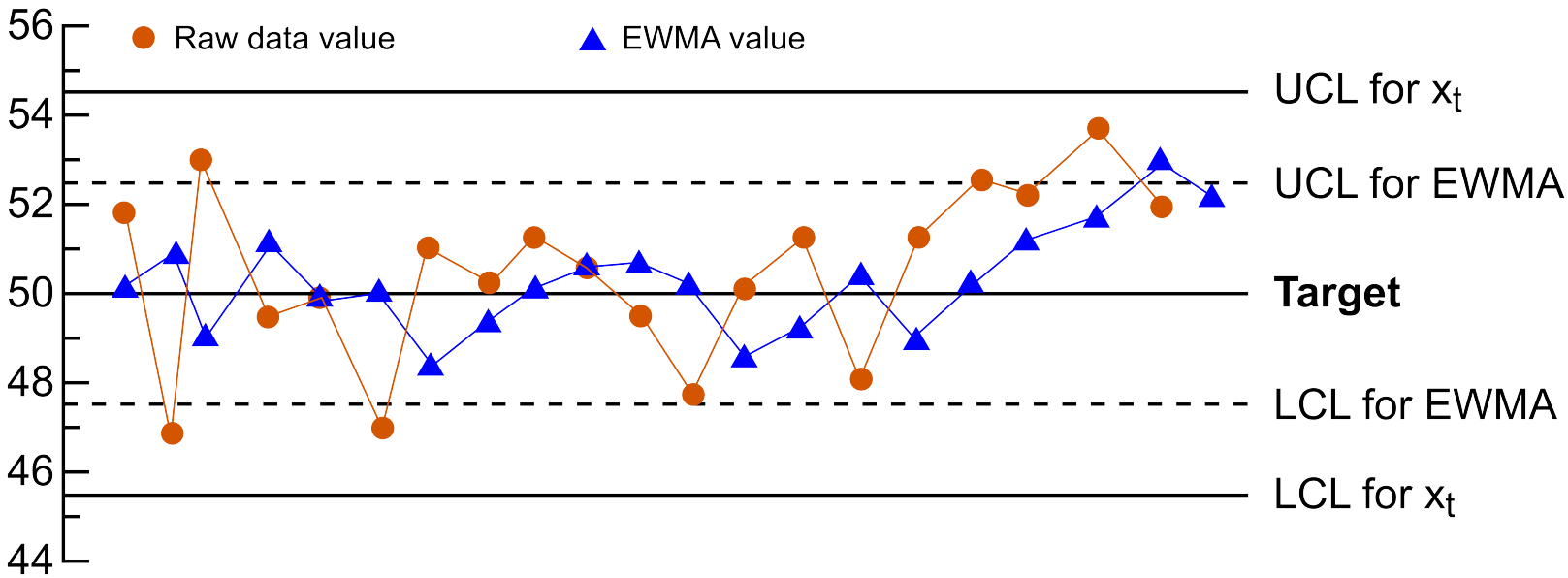
- Shows that EWMA is a one-step ahead predictor for $\hat{x}_{t+1}$
  1. x_t: current point, weighted by λ
  2. $\hat{x}_t$: historical data, weighted by $(1 - \lambda)$

EWMA example



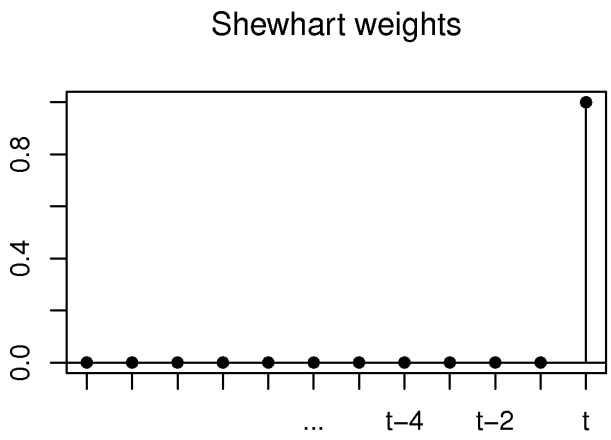
From: Hunter, "The Exponentially Weighted Moving Average", *Journal of Quality Technology*, **18**, p 203-210, 1986.

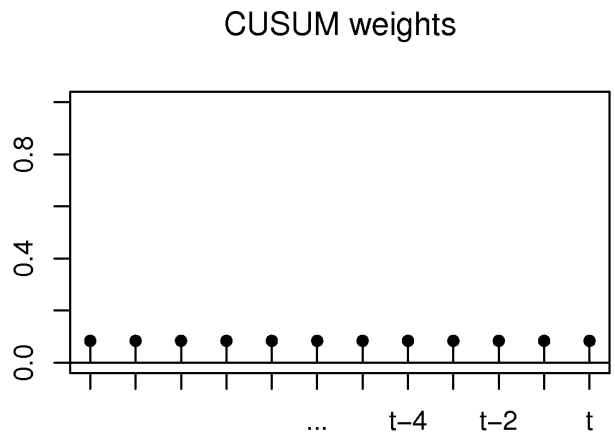
EWMA example

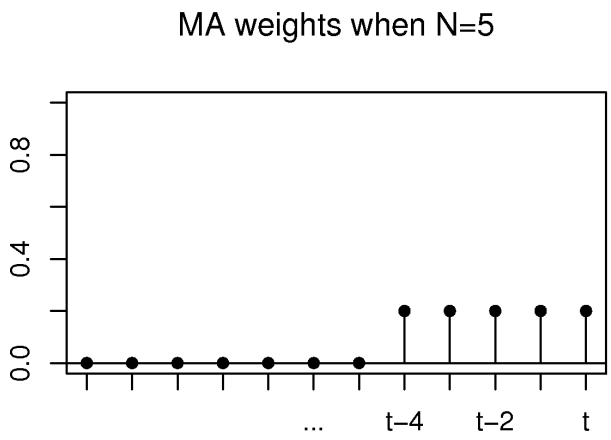


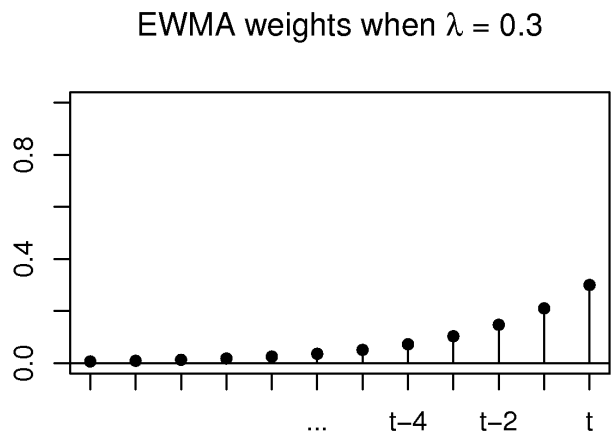
From: Hunter, "The Exponentially Weighted Moving Average", *Journal of Quality Technology*, **18**, p 203-210, 1986.

EWMA derivation example









- As $\lambda \to 0$: smoother chart, uses more history, less current data
- As $\lambda \to 1$: chart uses more current data (Shewhart-like)

EWMA derivation example

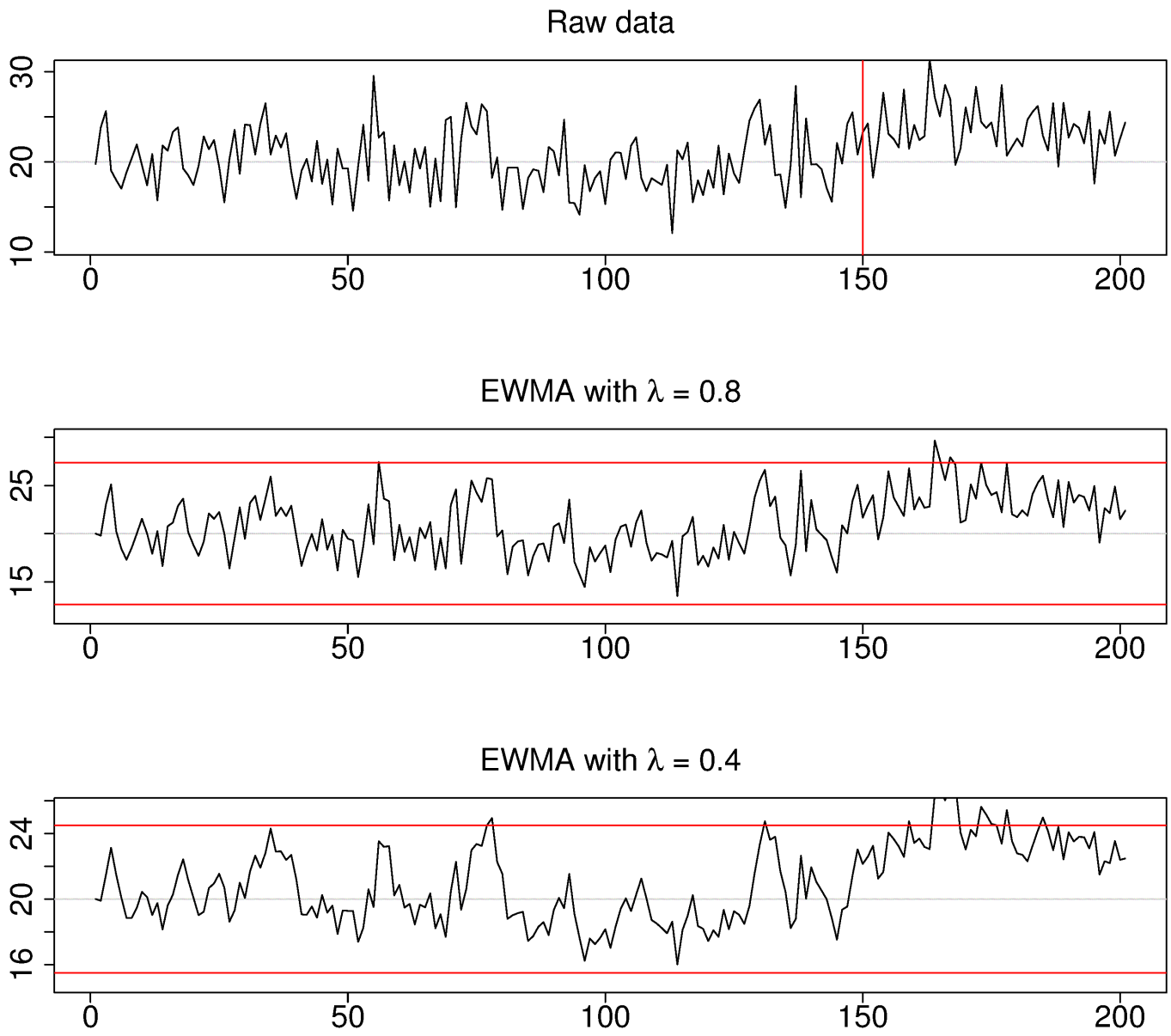
An alternative interpretation of the EWMA:

	$\hat{x}_{t+1} =$	$\lambda x_t +$	$\lambda(1-\lambda)x_{t-1} +$	$\lambda(1-\lambda)^2 x_{t-2} +$	$\lambda(1-\lambda)^3 x_{t-3} + \ldots$
$\lambda = 0.6:$	$\hat{x}_{t+1} =$	$0.6x_t +$	$0.24x_{t-1} +$	$0.096x_{t-2} +$	$0.0384x_{t-3} + \ldots$
$\lambda = 0.2:$	$\hat{x}_{t+1} =$	$0.2x_t +$	$0.16x_{t-1} +$	$0.128x_{t-2} +$	$0.1024x_{t-3} + \ldots$
MA chart*	$\hat{x}_{t+1} =$	$0.25x_t +$	$0.25x_{t-1} +$	$0.25x_{t-2} +$	$0.25x_{t-3}$

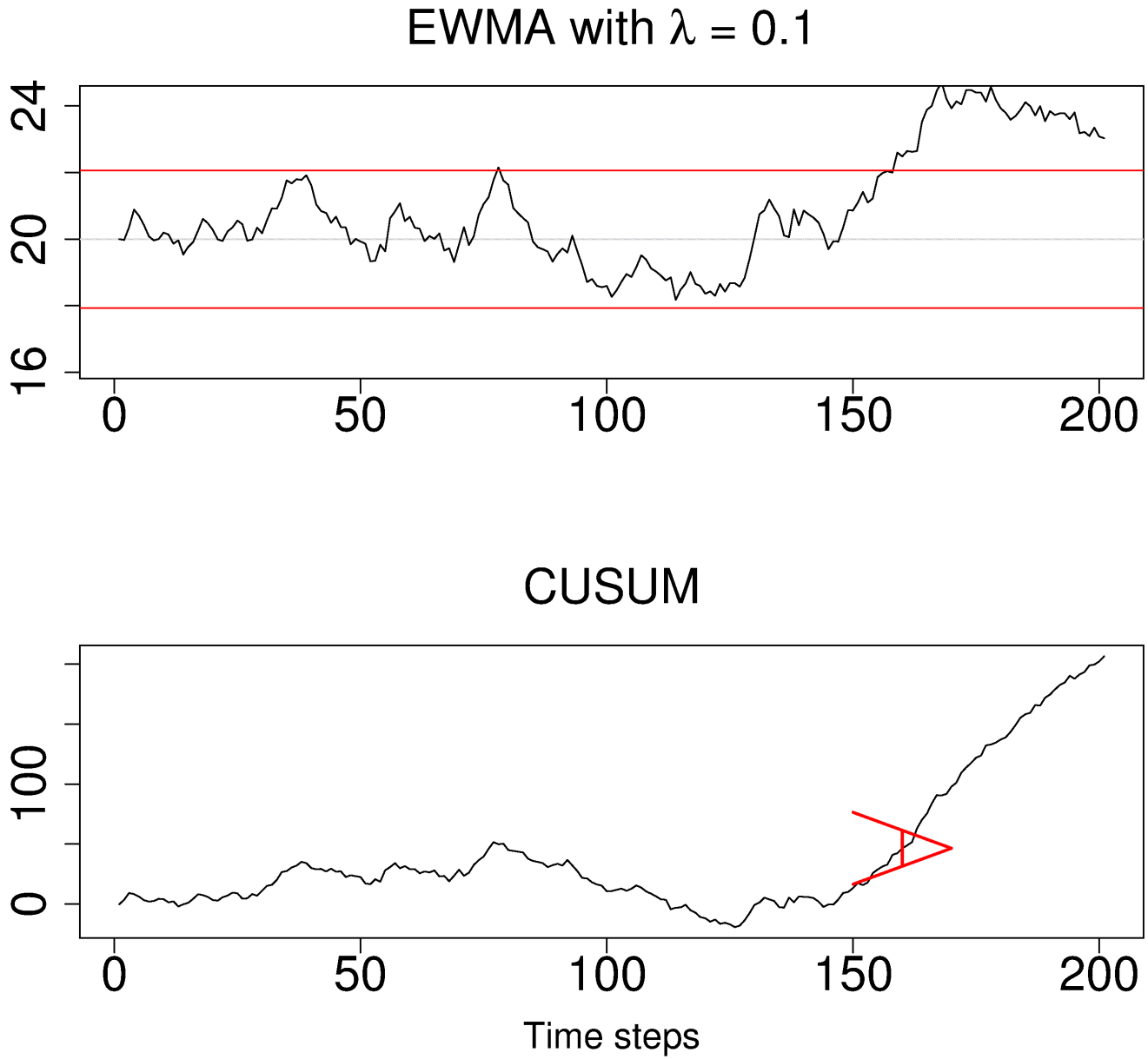
* MA chart with 4 samples per group: weight equals $1/4 = 0.25$

- As $\lambda \to 0$: smoother chart, uses more history, less current data
- As $\lambda \to 1$: chart uses more current data (Shewhart-like)

EWMA visual example



EWMA visual example



EWMA limits

$$\begin{aligned} \text{LCL} &= \bar{\bar{x}} - 3 \cdot \sigma_{\text{Shewhart}} \sqrt{\frac{\lambda}{2-\lambda}} \\ \text{UCL} &= \bar{\bar{x}} + 3 \cdot \sigma_{\text{Shewhart}} \sqrt{\frac{\lambda}{2-\lambda}} \end{aligned}$$

- σ_{Shewhart}: standard deviation from Shewhart chart $= \dfrac{\bar{S}}{a_n \sqrt{n}}$

Nice implementation: show both Shewhart and EWMA on the same chart

- This gives you the usual Shewhart monitoring, but for a
- slow-moving processes with long gaps between samples,
- the EWMA helpfully gives a one-step ahead prediction

Other charts

- The *S chart*: monitor variance
- The *R chart*: precursor to the *S chart* (not common anymore)
- *np chart* and *p chart*: monitoring proportions of pass/fail or good/bad ratings
- *Exponentially weight moving variance* (EWMV): used for monitoring product variability

What should we monitor?

Recall the aim is to **react early** to bad, or unusual operation:

- implies monitoring variables in near real-time
- laboratory measurements are good, but take longer to acquire
- don't wait for bad production to be over, catch it early

Key points

- Apply monitoring at every step in the manufacturing line/system
- Obtain low variability early on; don't wait to the end

Problem isn't how to monitor, rather, what do we monitor?

What should we monitor?

Measurements from real-time systems are:

- available more frequently (less delay) than lab measurements
- often are more precise
- more meaningful to the operating staff
- contains "fingerprint" of problem (helps for diagnosis)

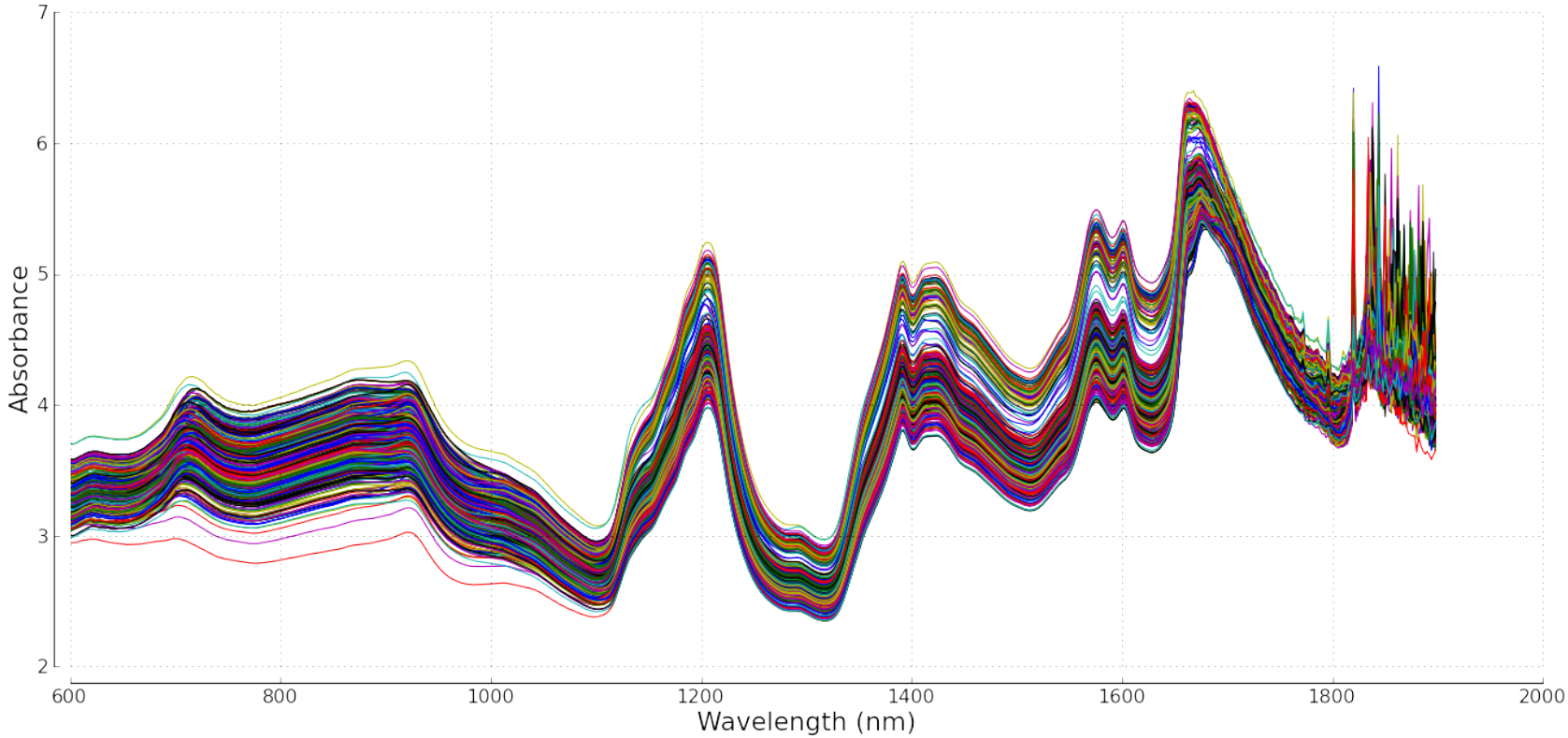
Variables don't need to be from on-line sensors: could also be a calculation

Lab measurements have long time delay:

- process already shifted by the time lab values detect a problem
- harder to find cause-and-effect for diagnosis

Monitoring in today's context

We don't measure a single number in many cases:



460 spectra (lines) measured at 650 evenly spaced wavelengths (x-axis). The y-axis is the absorbance at each wavelength.

What can we monitor with this data?

Monitoring with image data

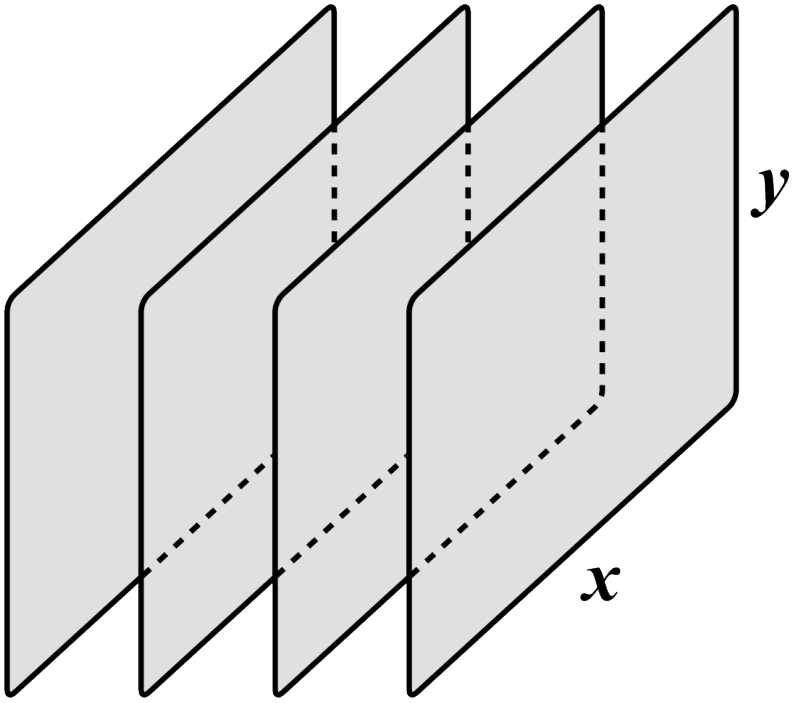


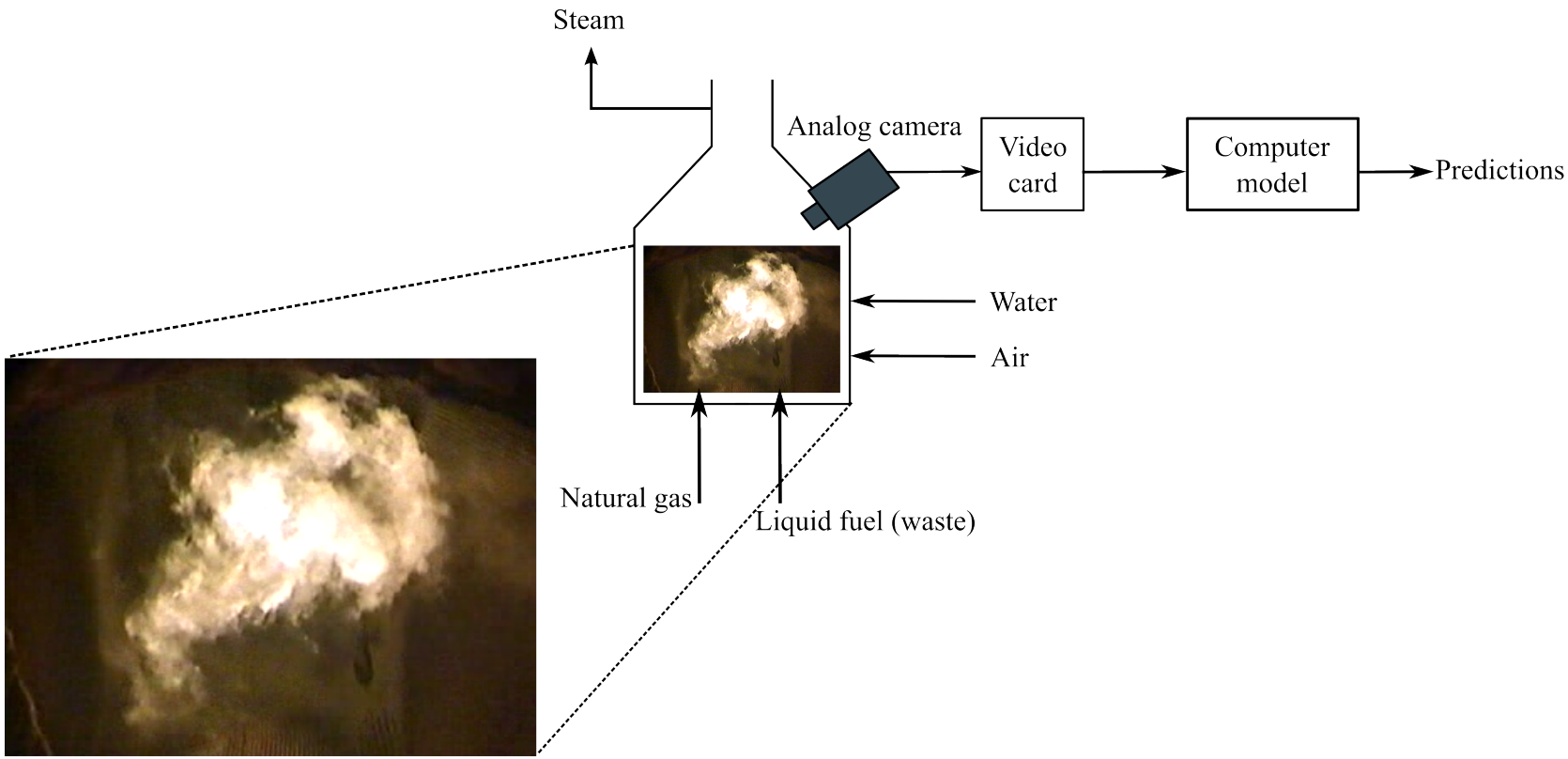
Image data: very common situation now

- many easy measurements for liquids and gases;
- but for solids: use image data
- medical imaging

The wavelength dimension (spectral dimension):

- 1 channel: grayscale
- 3 channels: colour image, RGB image
- multiple channels: hyperspectral image (e.g. NIR camera)

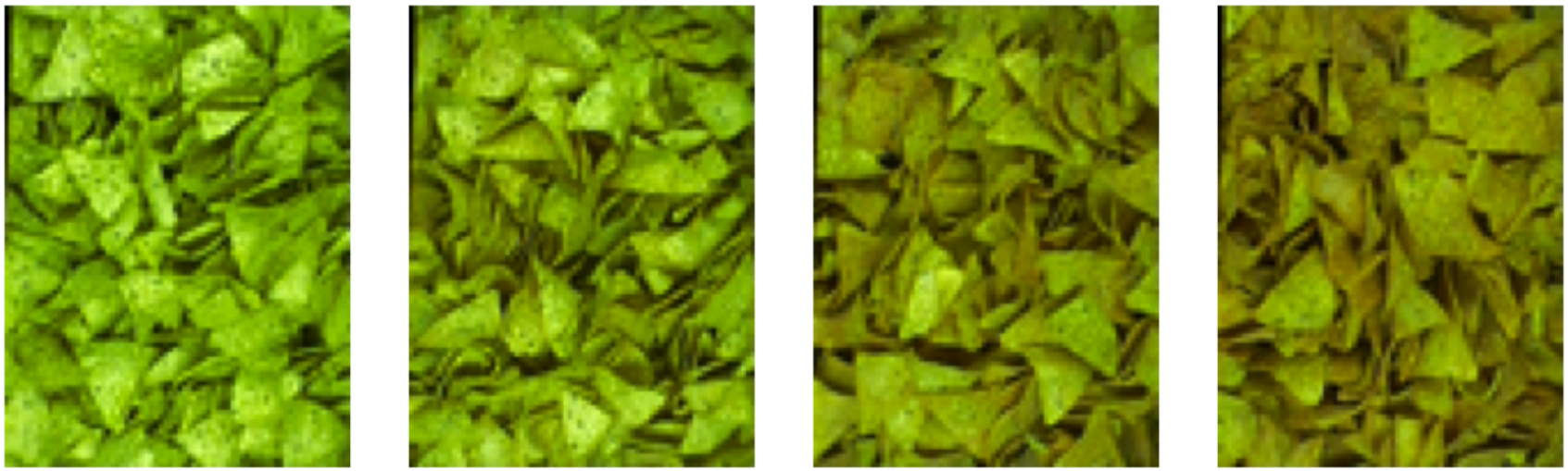
Flame monitoring application



- Liquid waste stream has variable energy content
- How can we maintain a desired steam flow from the boiler when the heating source is varying?

Seasoning application

We can see a change in product appearance with more seasoning: [1]

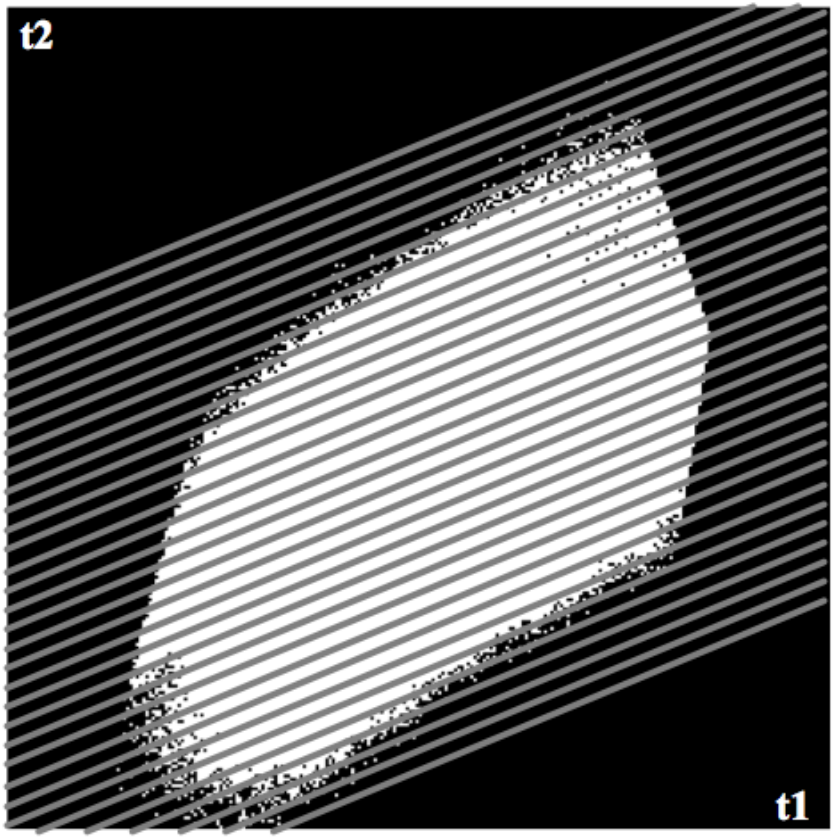


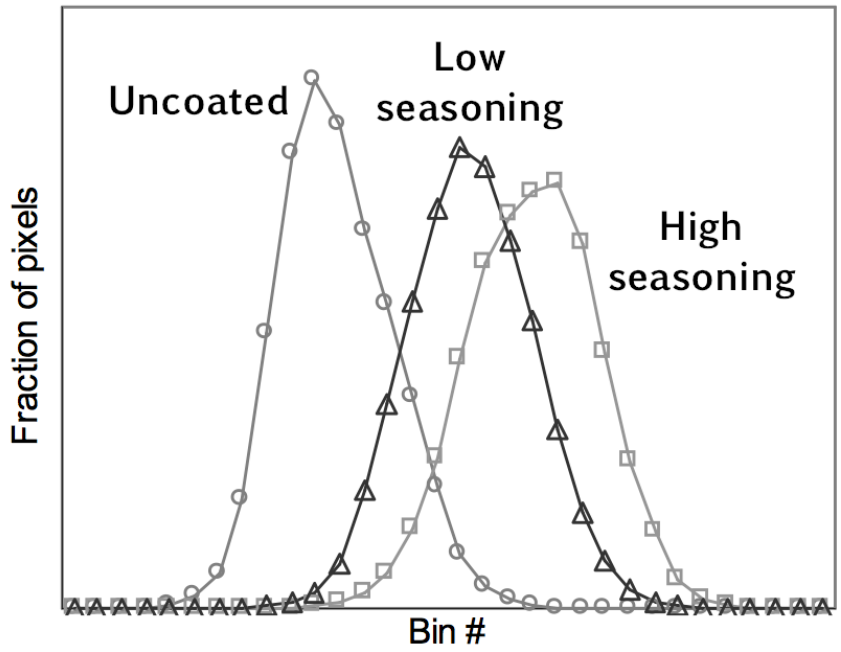
Increasing seasoning from left to right $\longrightarrow$

[1] From Honglu Yu's PhD (used with permission)

Feature extraction in the score space (phase 1)

Create bins in the score space (eigen-decomposition of the raw image data); then count pixels:





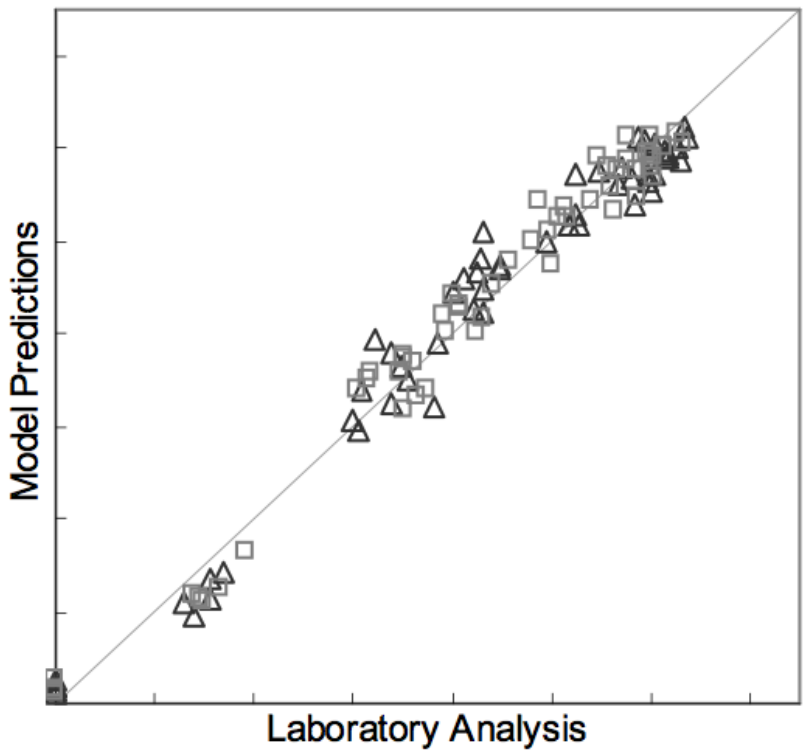
Direction of bins: direction in which seasoning level shifts the score plot [2]

[2] Other binning methods described in journal publications and thesis

Model building and predictions (phase 1)

PLS predictive model:

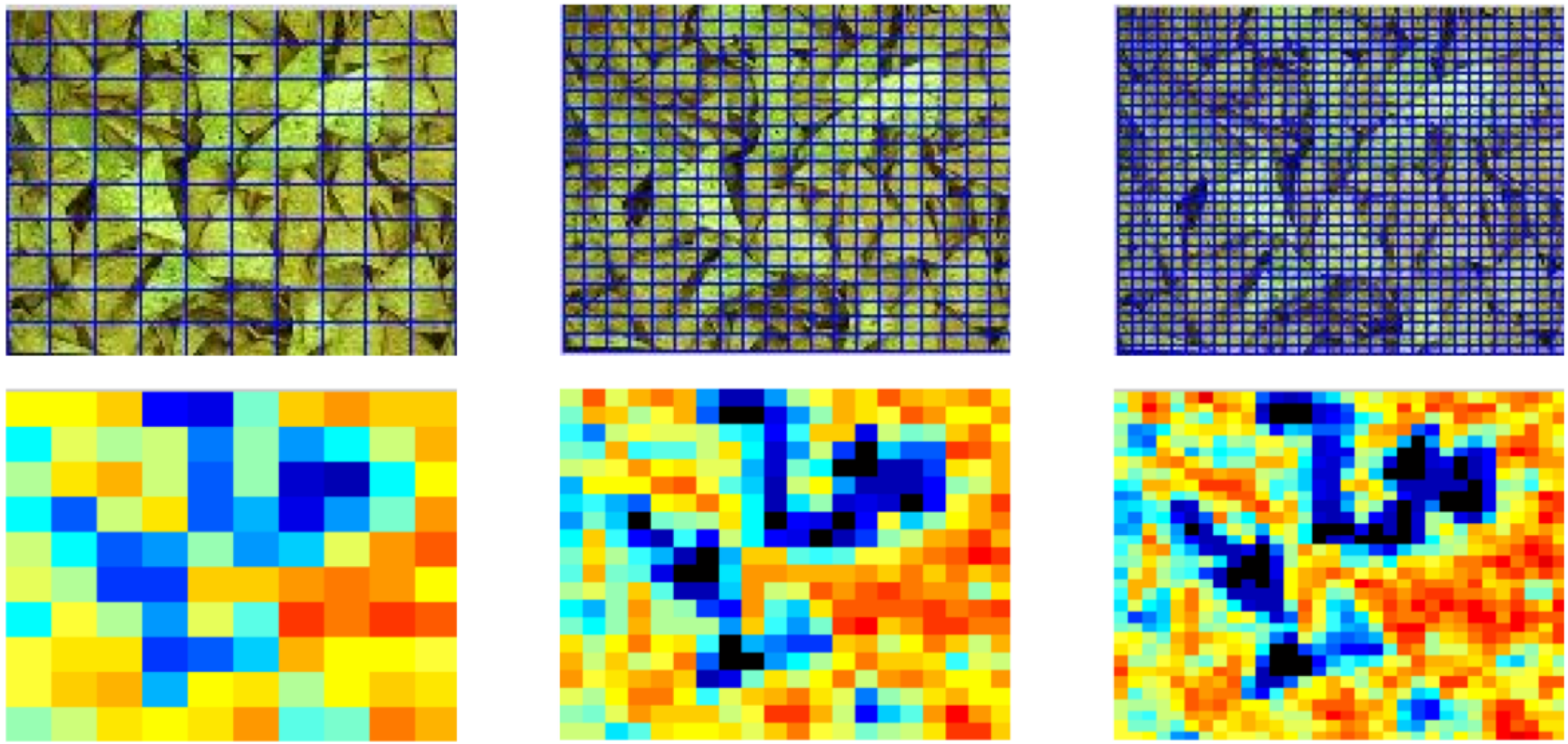
- **X**-space: cumulative histogram of *fractional* bin counts
  - Note the **X**-space will be very collinear
- y-variable: seasoning level



triangles = training data; squares = testing data

Apply the predictive model to new images (phase 2)

- Apply procedure to subset of pixels:



Monitoring system built for:

- average seasoning level (also placed under automatic feedback control)
- seasoning *variance*

Monitoring in industrial practice

- Widely used in industry, at all levels
- Management: monitor plants, geographic region, countries (e.g. hourly sales by region)
  - Dashboards, ERP, BI, KPI
- Challenges for you:
  - Getting the data out
  - Real-time use of the data (value of data decays exponentially)
  - Training is time consuming
  - Bandwidth/network/storage

General workflow I

1. Identify variable(s) to monitor.
2. Retrieve historical data (computer systems, or lab data, or paper records)
3. Import data and just plot it.
   - Any time trends, outliers, spikes, missing data gaps?
4. Locate regions of stable, common-cause operation.
   - Remove spikes and outliers
   - Cleaned data is your phase 1 data
5. Split phase 1 data into a 60% and 40% split.
   - The 60% split is for calculating model limits
   - The 40% is for testing later on.
6. Keep outlier data as a separate testing set: to validate detection
7. Calculate control limits (UCL, LCL), using formula, using 60% data chunk

General workflow II

8. Test your chart on **new, unused** data. *How does my chart work?*
   - Quantify type I error on cleaned 40% chunk
   - Quantify type II error on outlier data
9. Adjust the limits
10. Repeat these steps, as needed to achieve levels of error
11. Run chart on your desktop computer for a couple of days
    - Confirm unusual events with operators; would they have reacted to it? False alarm?
    - Refine your limits
12. Not an expert system - will not diagnose problems:
    - use your engineering judgement; look at patterns; knowledge of other process events; *troubleshooting* skills!
13. Demonstrate to your colleagues and manager
    - But go with dollar values.
14. Installation and operator training will take time

General workflow III

15. Listen to your operators
    - make plots interactive - click on unusual point, it drills-down to give more context

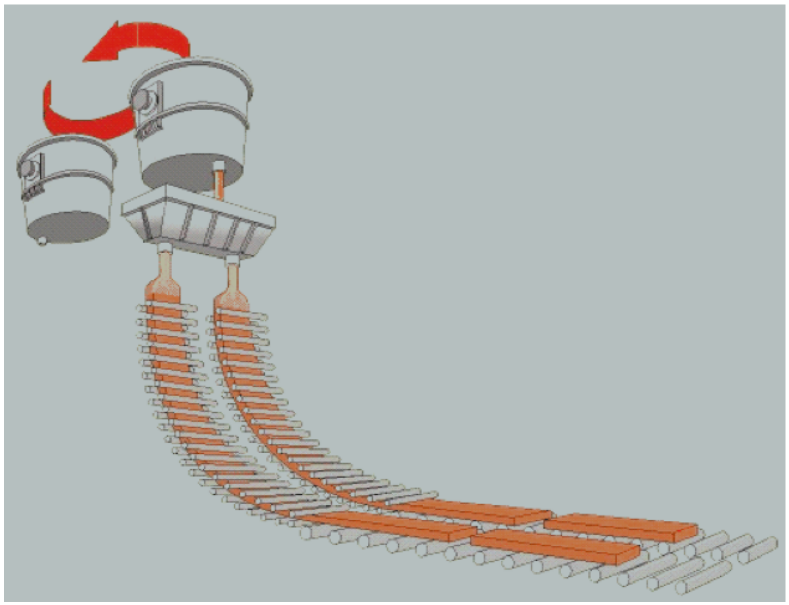
Industrial case study: Dofasco

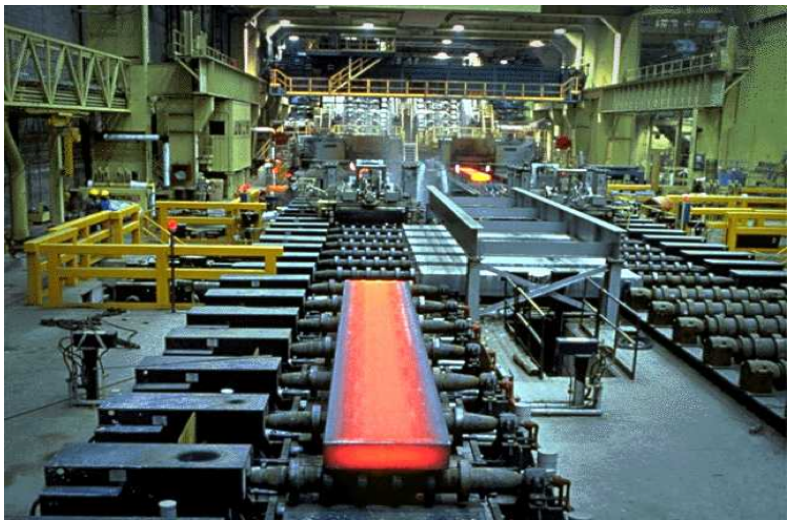
- ArcelorMittal in Hamilton (formerly called Dofasco) has used multivariate process monitoring tools since 1990's
- Over 100 applications used daily
- Most well known is their casting monitoring application, Caster SOS (Stable Operation Supervisor)
- It is a multivariate monitoring system

Dofasco case study: slabs of steel

All screenshots with permission of Dr. John MacGregor

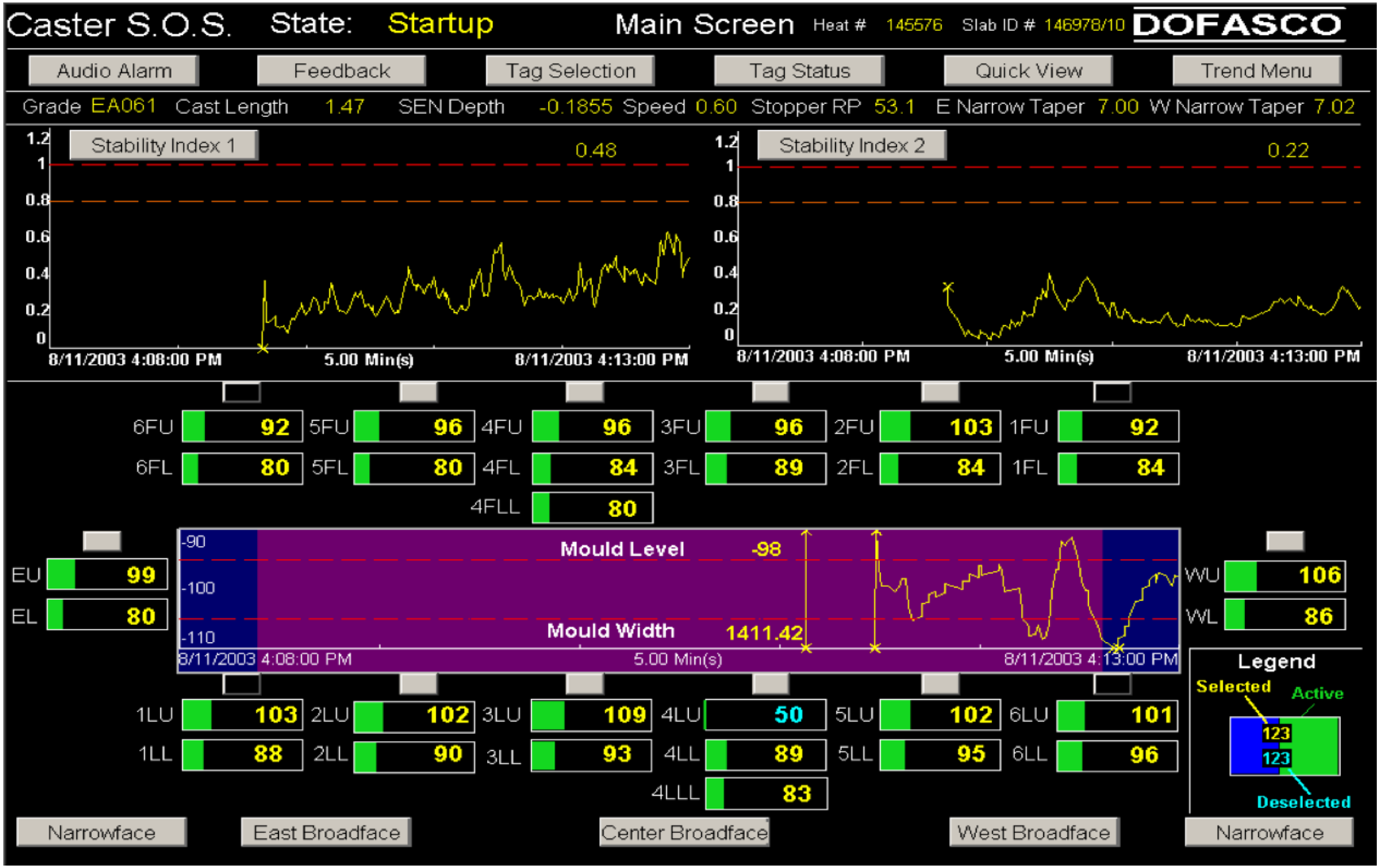
Dofasco case study: casting



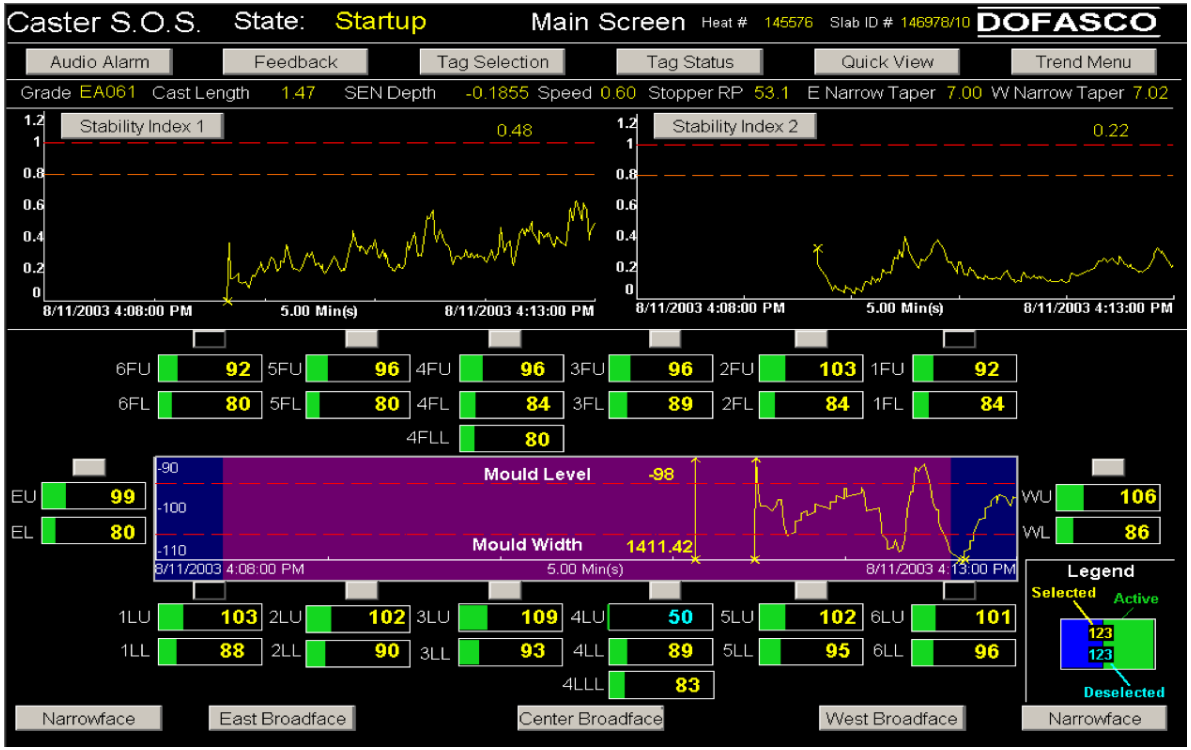


Dofasco case study: breakout

Dofasco case study: monitoring for breakouts

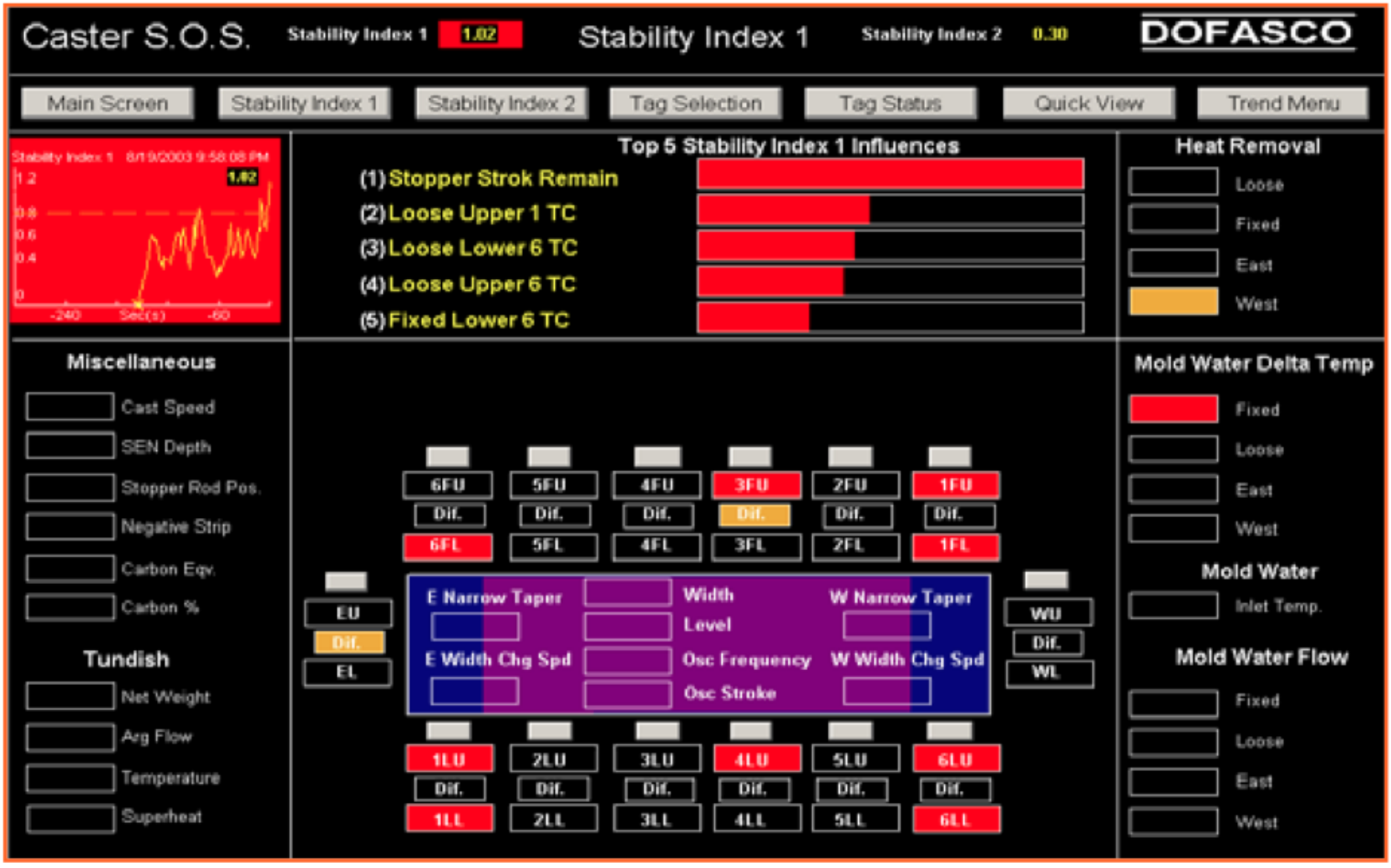


Dofasco case study: monitoring for breakouts

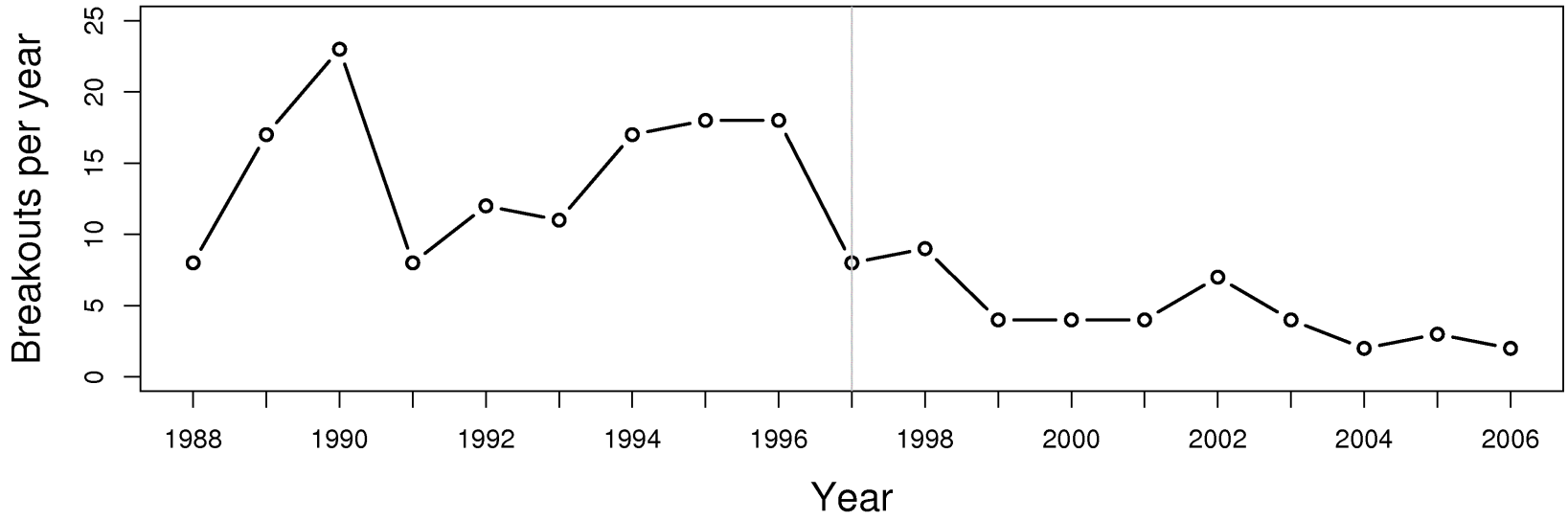


- Stability Index 1 and 2: one-sided monitoring chart
- Warning limits and the action limits.
- A two-sided chart in the middle
- Plenty of other operator-relevant information

Dofasco case study: an alarm



Dofasco case study: economics of monitoring



- Implemented system in 1997; multiple upgrades since then
- Economic savings: more than $ 1 million/year
  - each breakout costs around $200,000 to $500,000
  - process shutdowns and/or equipment damage
  - more than justifies the costs and person-hours to implement the system

Process capability: centered process

Process capability ratio (PCR) can be calculated for any attribute.

$$\text{PCR} = \frac{\text{Upper specification limit} - \text{Lower specification limit}}{6\sigma}$$

- Use an estimate for σ
- LSL is not LCL; and USL is not UCL from Shewhart chart
- LSL and USL are set by customers, or internal criteria

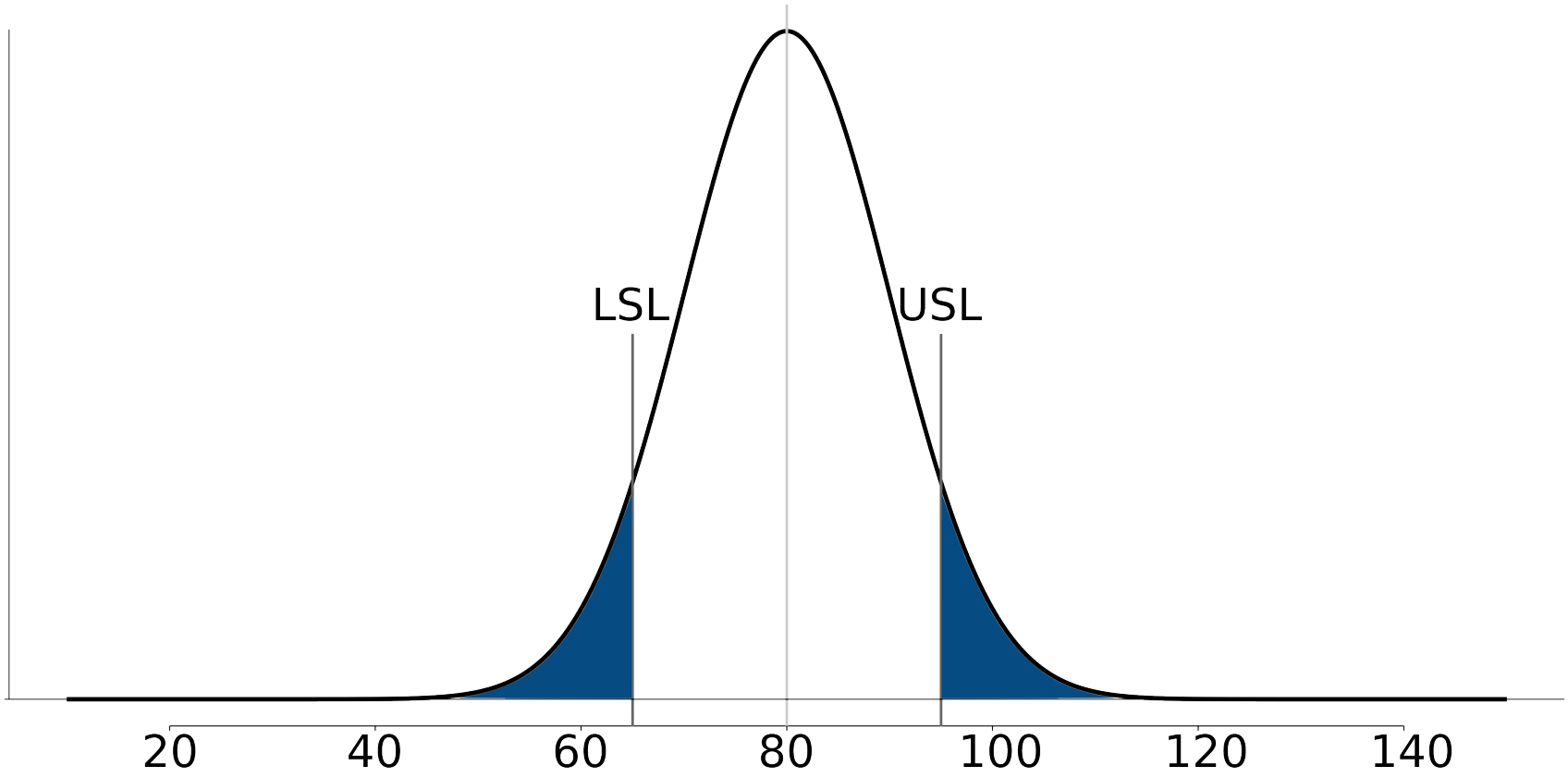
Strong assumptions used for PCR:

- assumes the attribute has normal distribution (check with `qqPlot`)
- assumes centered system between LSL and USL
- assumes PCR calculated when process was stable

PCR interpretation: process "width"

Let mean=80, LSL=65, USL=95 and $\hat{\sigma} = 10$

Implies: PCR $= 0.5$

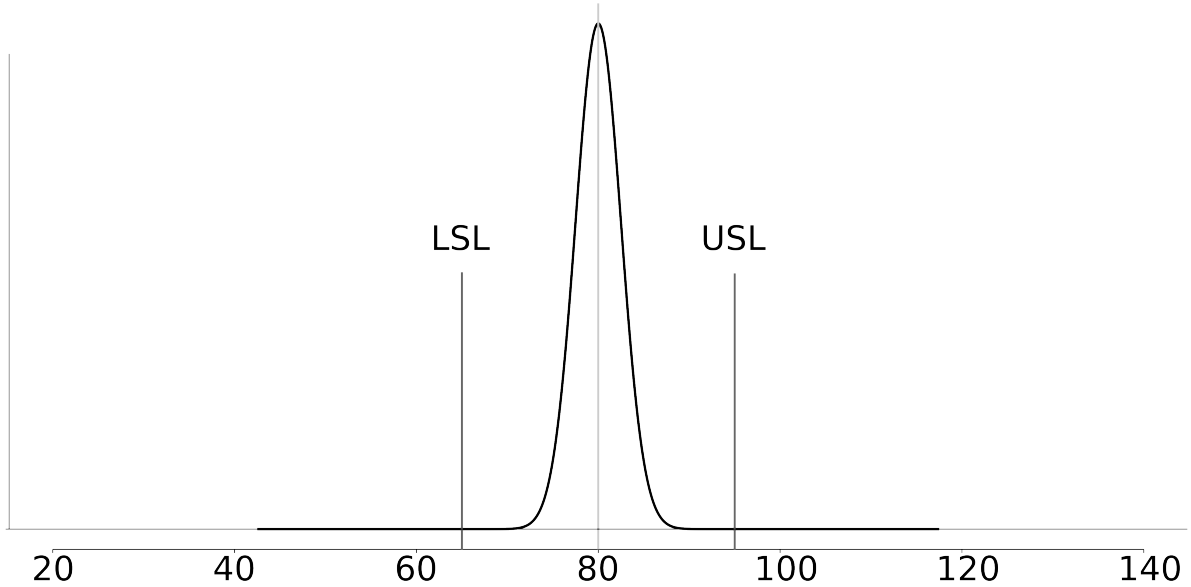


- z for LSL $= (65 - 80)/10 = -1.5$
- z for USL $= (95 - 80)/10 = +1.5$
- Shaded area probability $=$ `pnorm(-1.5) + (1-pnorm(1.5))` $= 13.4\%$

PCR interpretation: process "width"

Let mean=80, LSL=65, USL=95 and $\hat{\sigma} = 10/4 = 2.5$

Implies: PCR $= 2.0$



- z for LSL $= (65 - 80)/2.5 = -6$
- z for USL $= (95 - 80)/2.5 = +6$
- Shaded area probability = about 0 $(1.973 \times 10^{-9} \times 100)$
- Process width: 12σ
- Why is it called process capability?
  - If you cannot make your variance (standard deviation) smaller, your process is not capable

PCR interpretation: uncentered process

- Processes not often centered between LSL and USL

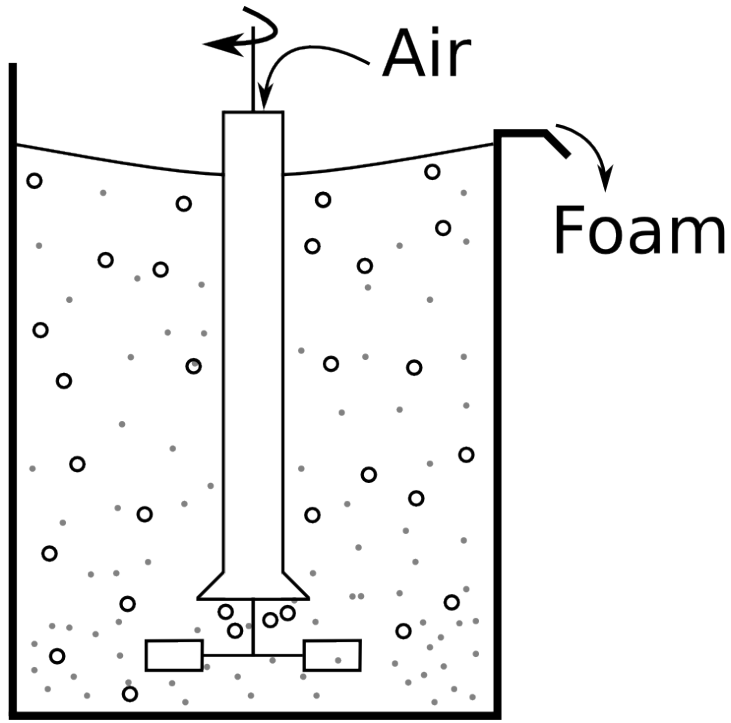
$$\text{PCR}_k = \text{C}_{pk} = \min\left(\frac{\text{Upper specification limit} - \bar{\bar{x}}}{3\sigma}; \frac{\bar{\bar{x}} - \text{Lower specification limit}}{3\sigma}\right)$$

- $\bar{\bar{x}}$ from Shewhart chart
- One-sided limit: taken on the worst side!
- Cpk $= 1.3$: minimum requirement
- Cpk $= 1.7$: requested for safety and other critical applications.
- Cpk $= 2.0$: termed a 6-sigma process: it can move 6σ units left or right

Note: Cpk and Cp are only useful for a process which is stable

Example

A tank uses small air bubbles to keep solid particles in suspension. If too much air is blown into the tank, then excessive foaming and loss of valuable solid product occurs; if too little air is blown into the tank the particles sink and drop out of suspension.



Which monitoring chart would you use to ensure the airflow is always near target?

Example

Describe how a monitoring chart could be used to prevent over-control of a batch-to-batch process. (A batch-to-batch process is one where a batch of materials is processed, followed by another batch, and so on).

Example

Final exam, 2010
The most recent estimate of the process capability ratio for a key quality variable was 1.30, and the average quality value was 64.0. Your process operates closer to the lower specification limit of 56.0. The upper specification limit is 93.0.

What are the two parameters of the system you could adjust, and by how much, to achieve a capability ratio of 1.67, required by recent safety regulations. Assume you can adjust these parameters independently.

Example

The following values are the particle size of the most recent 20 shipments from a supplier, taken from their certificates of analysis:
`50.9, 52.9, 51.6, 50.8, 54.6, 52.9, 53.1`
`48.4, 51.6, 53.1, 53.8, 52.4, 53.1, 50.8`
`54.6, 52.9, 50.0, 53.8, 54.6, 52.2`

Calculate the supplier's capability, given their lower specification limit of $45\mu m$ and their upper limit at $59\mu m$.

Clearly state all assumptions you make during the calculations (median $= 52.9\mu m$, sd $= 1.64\mu m$)

Example

Plastic sheets are manufactured on your blown film line. The Cp value is 1.7. You sell the plastic sheets to your customers with specification of 2 mm $\pm$ 0.4 mm.

- List three important assumptions you must make to interpret the Cp value.
- What is the theoretical process standard deviation, σ?
- What would be the Shewhart chart limits for this system using subgroups of size $n = 4$?
- Illustrate your answer from part 2 and 3 of this question on a diagram of the normal distribution.

Answer