

-自动化学院学科核心课-

# 检测技术与自动化

## 第6章 现代检测技术-5-软测量技术





## 什么是软测量：

所谓软测量，就是应用计算机技术，以软件代替硬件（传感器）的功能，来实现对某参数的在线检测。



完全不需要传感器么？

## 软测量的意义：

许多被控对象属于**复杂系统**，具有非线性、时变性、不确定性等特性，很难建模，许多人开始研究大规模复杂系统的性质、优化和控制问题，使复杂系统研究成为控制技术发展的一个重要领域。

为了使生产或实验达到某工艺设计要求，常要求对一些**生化参数**进行控制，例如对啤酒发酵时的糖度进行控制，对谷氨酸发酵过程中的浅糖浓度进行控制等等。这时，有没有**自动检测**这些生化参数的仪表就成了能否构成**自动控制系统**的关键。



血脂生化检测仪

解决分量、生化量在线自动检测问题有**两种方法**：

**1) 传统方法：**研发一种**有单独传感器和仪表电路**等硬件及相应软件，**具有很好选择性**（对被测量敏感，对其它量不敏感）的在线检测仪表。需要花很长的时间研究测量机理，开发材料、结构、硬件和软件，还需试验、修改与标定。

**2) 软测量方法：****充分利用**其它测量仪表已获得的**工艺参数**，再增加一些容易获取的参数，按照这些参数与被测变量之间存在关系（模型），通过计算**估计推断被测变量**。曾被国际著名控制专家T. J. McacVoy列为未来控制领域需要研究的几大方向之一。

## 为什么要采用软测量技术：

在现代工业生产过程中，存在一大类这样的变量：它们由于**经济或者技术原因**，目前尚难以或暂时无法通过传感器进行检测；但同时又是需要严格控制的，与产品质量相关的重要的过程参数。

采用软测量技术 (Soft sensing techniques)或软测量仪表 (Soft sensor)，就能解决这类变量的测量问题。





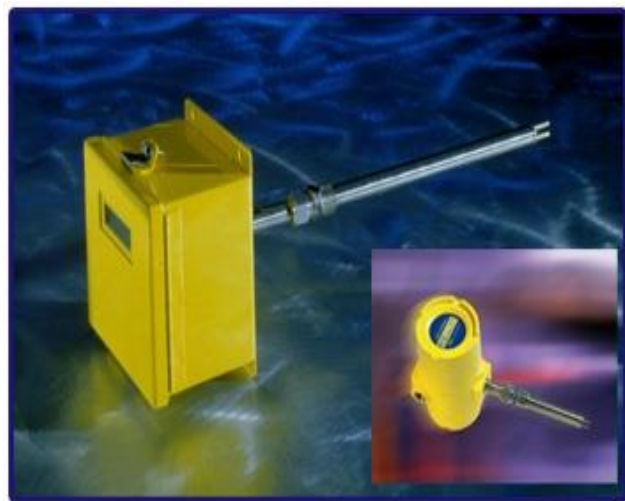


对难测量或暂时不能用单独的仪表进行测量的被测变量（**主导变量**），选一些容易测量的变量（**二次变量或辅助变量**），构成辅助变量与被测变量之间的**数学关系**，用数据处理软件代替单独的仪表这种硬件，通过推算确定被测变量。

软测量基本思想早在复合测量上就有所体现，例如**用差压法测量气体质量流量**。人们通过研究找到了差压、气体压力、气体温度和气体质量流量之间的数学关系：

$$Q_m = k_1 k_2 \sqrt{\rho_0 \frac{293.15 p}{1.029(T + 273.15)} \Delta p}$$

$$Q_m = k_1 k_2 \sqrt{\rho_0 \frac{293.15 p}{1.029(T + 273.15)} \Delta p}$$



气体质量流量计

将差压、气体压力和气体温度三个容易测量的变量作为辅助变量进行测量，然后按数学关系便可计算确定气体质量流量。

由于可直接用现有的仪表获取辅助变量，则不再需要研发专用仪表，投入少，周期短，维护简单，又可获得与单独专用仪表一样的快速响应及计量效果。

**臭氧浓度分析仪**可作为体现**软测量思想**的例子。

用温度传感器和压力传感器测量样气的**温度**和**压力**，用光强仪测量**样气对紫外光的吸收特性**，按以下确定的数学模型进行计算，可测得臭氧浓度。

$$C = -\frac{0.101325T}{273\alpha_0 P} 10^6 \ln(I / I_0)$$



**臭氧浓度分析仪**



以上两例都存在一个起主要作用的**辅助变量**，即气体质量流量测量的差压和臭氧浓度测量的紫外光通过强度，其它辅助变量主要用于环境因素的**自动补偿**，以提高测量精度。这种补偿按照已知的机理模型用软件计算自动进行，通过多传感器的复合测量实现环境因素影响的自动补偿。

上述两例与**软测量技术**存在相当的差别：有的资料将**间接测量**都算为软测量是不合适的，大多数间接测量建立的是辅助变量与被测变量之间一对一的关系，可以用模拟硬件直接实现。

## 【例】

两相流或多相流是指由两种或两种以上不同物质或成分构成的流动（包括油/气、油/水两相流，油/气/水多相流和气/固两相流等），其参数检测一直是一个难题。

国内外常采取的两相流检测技术大体可归为三类：

- 1) 采用传统的单相流仪表和两相流测试模型结合，其测量精度和使用条件有限；
- 2) 采用微波、核磁共振、过程层析成像等新型检测技术，其测量手段复杂且成本很高；
- 3) 采用**软测量技术**，可降低费用又可保证一定的精度。



**软测量**就是选择与被测变量（主导变量Primary Variable,如产品分布、物料成分等）相关的一组可测变量（辅助变量Secondary Variable，如温度、压力等）；构建某种以可测变量为输入，被测变量为输出的数学模型，使用计算机来进行模型的数值运算，从而得到被测变量的估计值的过程。

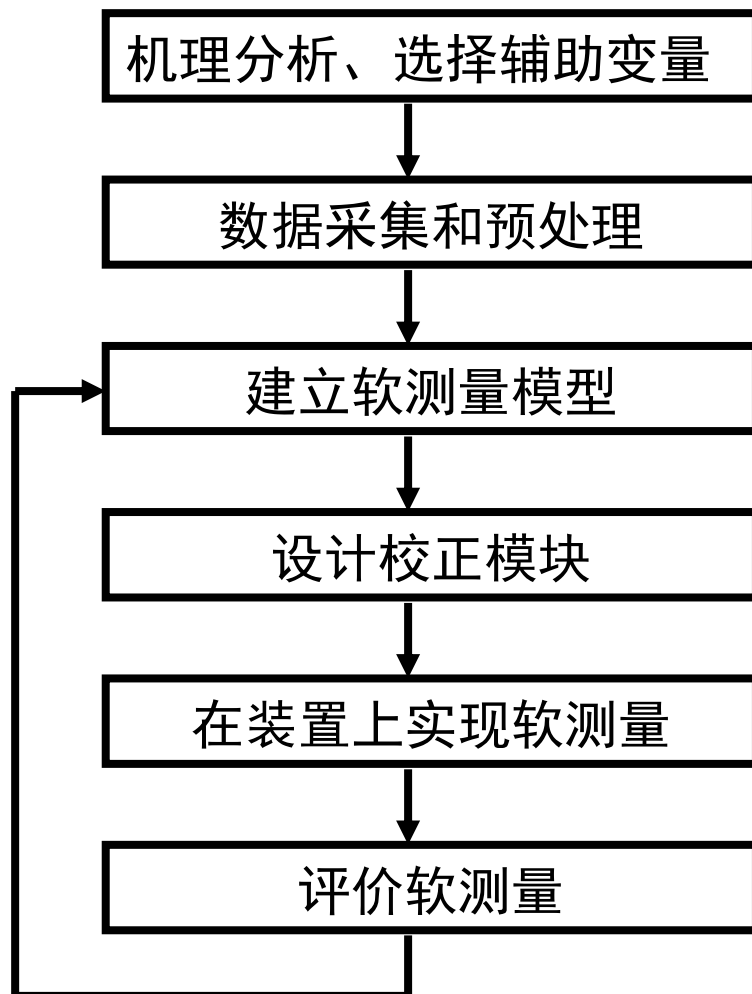
**软测量技术主要包括四个方面：**

- (1)辅助变量的选取
- (2)测量数据的处理
- (3)软测量模型的建立
- (4)软测量模型的在线校正



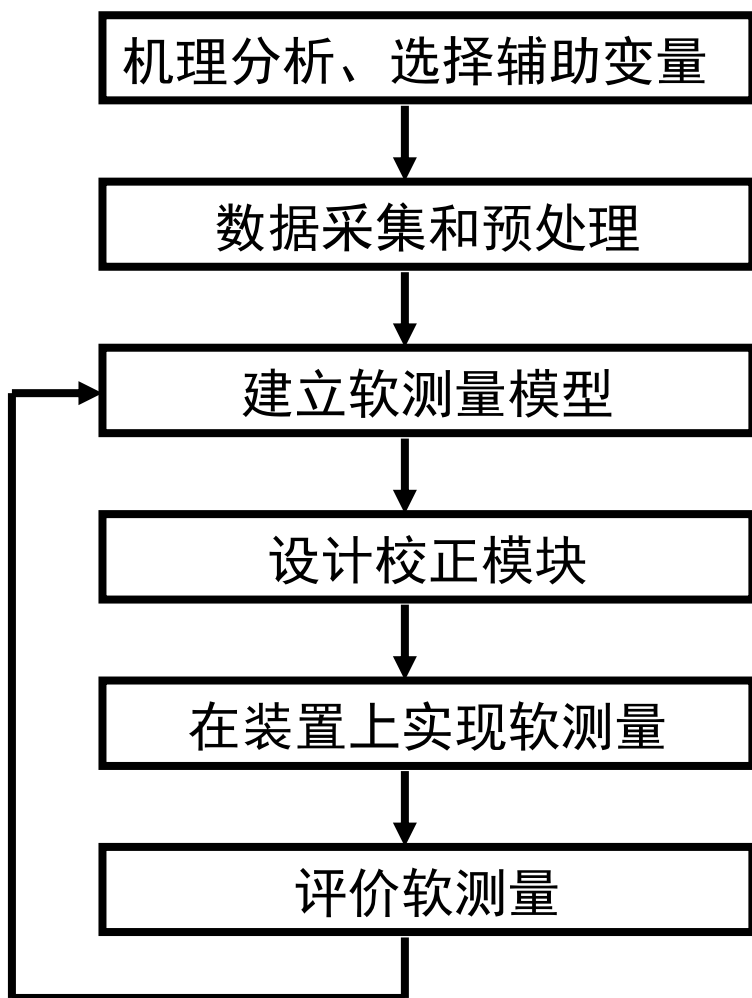
## 软测量的设计步骤：

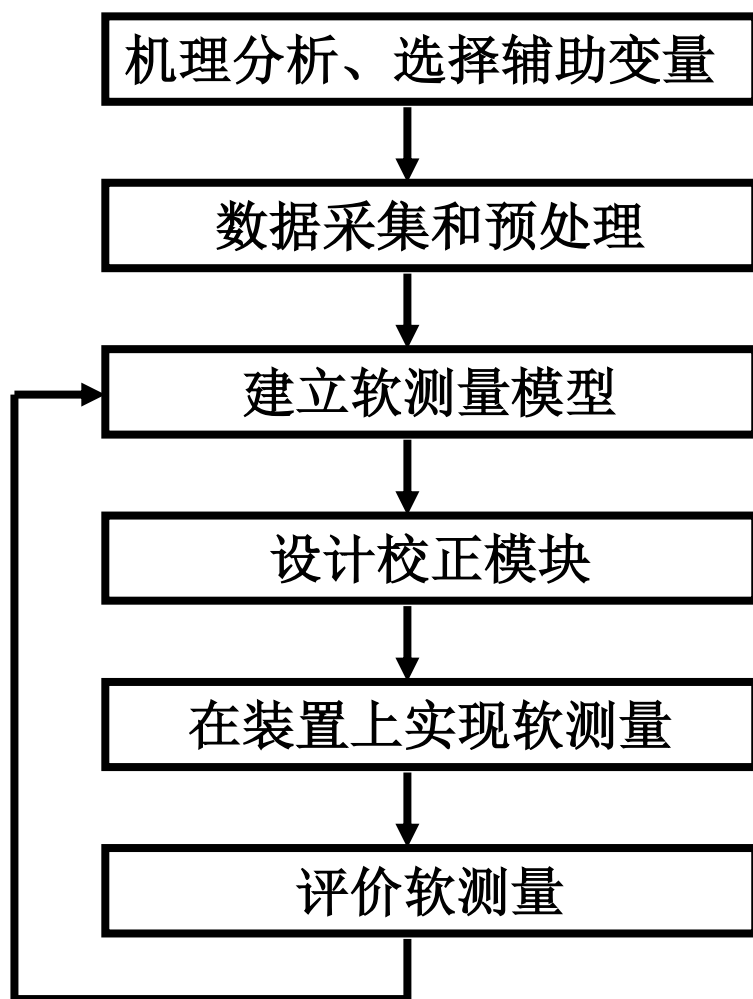
软  
测  
量  
开  
发  
流  
程





1. 了解工艺流程和控制系统
2. 明确软测量任务，确定主导变量
3. 分析变量的可观、可控性
4. 初步选择辅助变量





1. 采集辅助变量的测量数据和主导变量的化验数据

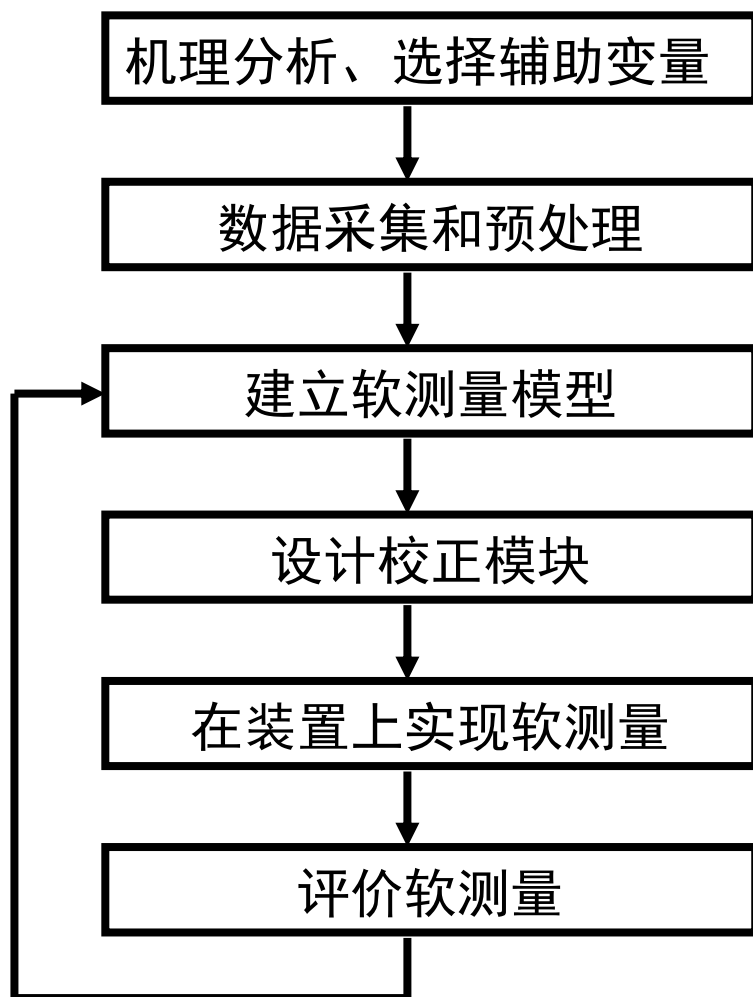
2. 数据校正和数据变换

3. 数据相关性分析

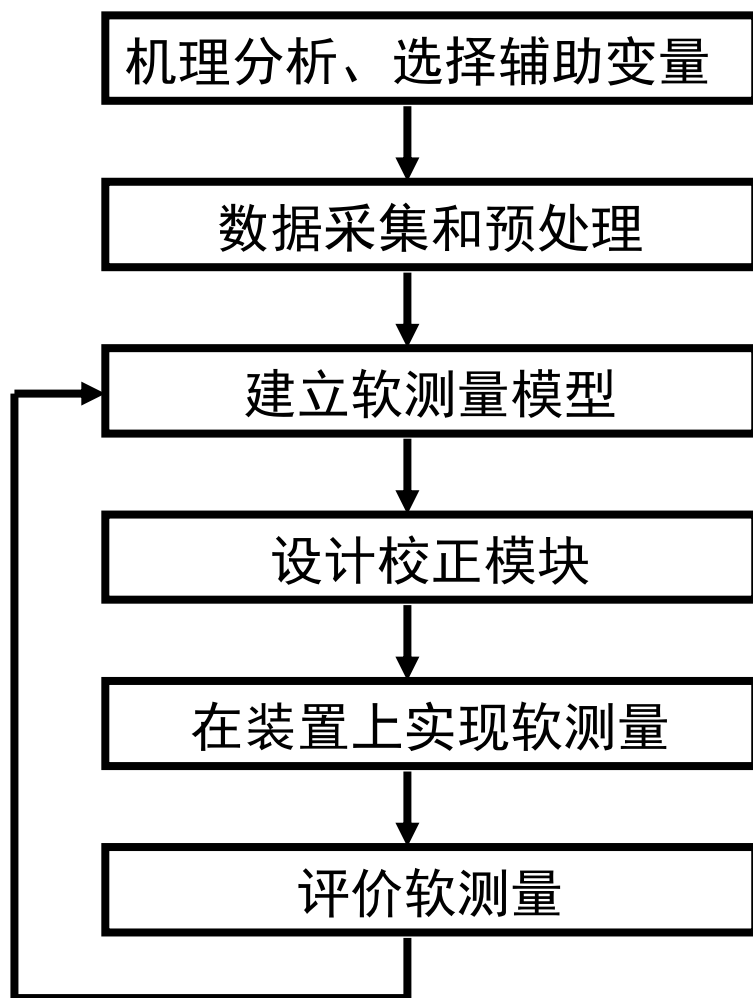
4. 主元分析

与建模有关

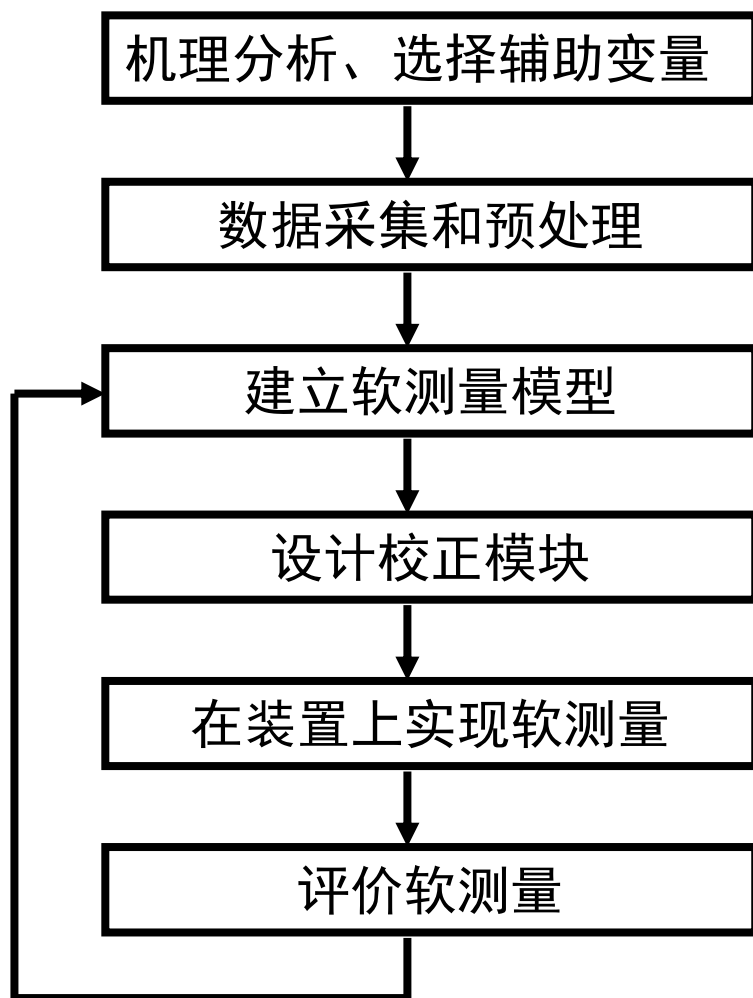




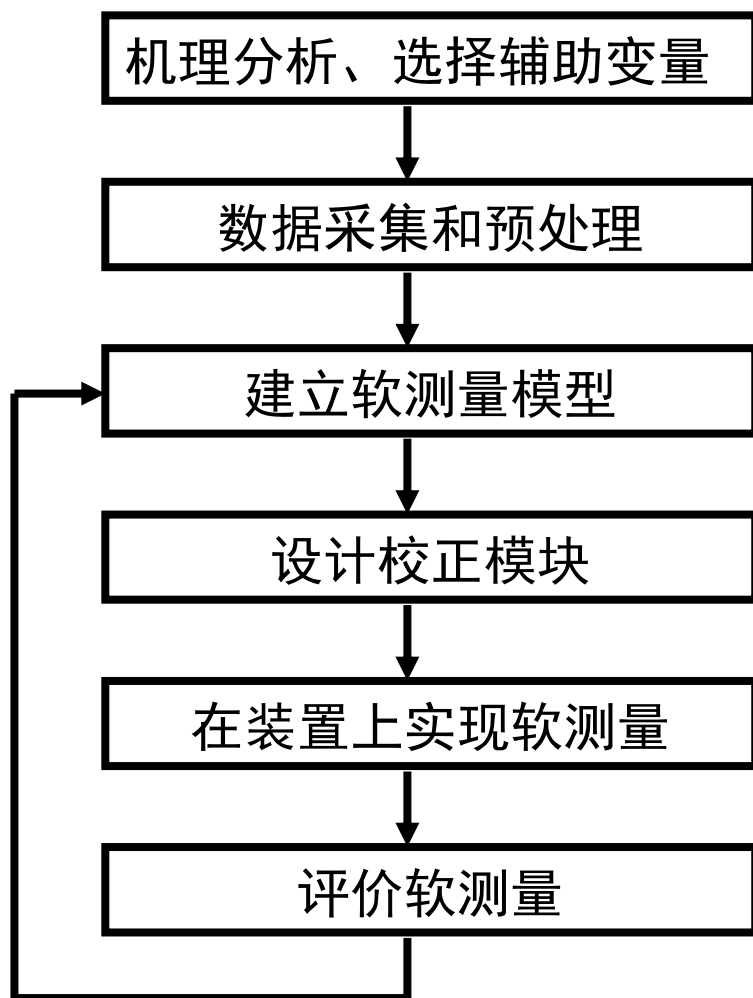
- 1. 确定辅助变量
- 2. 选择软测量模型结构
- 3. 确定软测量模型系数



- 1. 选择短期校正方法（在线）
- 2. 选择长期校正方法（在线或离线）



1. 设计在DCS上实现软测量的数据处理、模型计算和在线校正模块；
2. 设计工艺员修改参数界面和操作人员观测、输入化验值界面；
3. 设计模型报警模块。



- 1. 是否满足精度要求
- 2. 如不满足，修改软测量模型



## 5. 软测量技术

5.1 辅助变量的选取 \*

5.2 测量数据的处理

5.3 软测量模型的建立 \*

5.4 软仪表的在线校正

5.5 软测量的工业应用



## 5.1 辅助变量的选取

### 1. 选择影响主导变量的可测相关变量作为辅助变量

选择的方法往往从间接质量指标出发

➤例：估计精馏塔塔顶产品的成分软测量

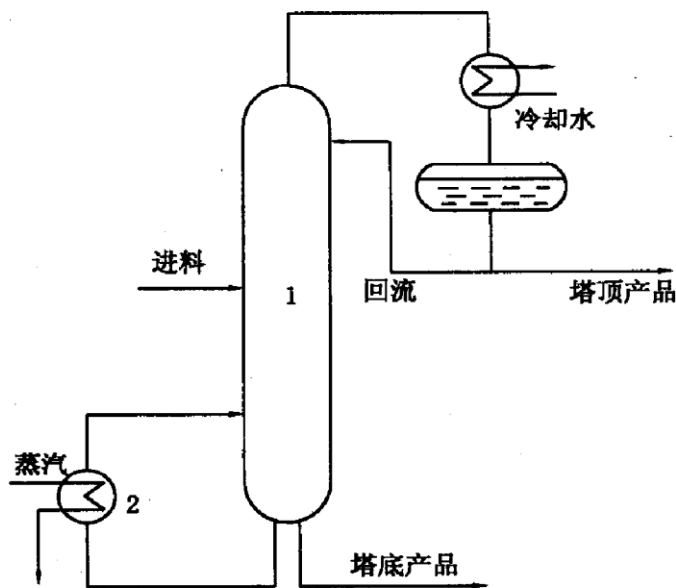


图 6-3 精馏过程示意图

1-精馏塔；2-蒸汽加热器

#### ■选择初始辅助变量

- 塔的进料特性
- 塔的加热特性
- 塔顶回流特性
- 塔顶操作状态
- 塔抽出料特性





## 2. 变量数据的选择

- 通过机理分析，从影响被估计变量的变量中选择响应灵敏、测量精度高的作为最终辅助变量（降维处理）。
- 如果缺乏机理知识，则可用回归分析的方法找出影响被估计变量的主要因素，这需要大量的观测数据，将原始辅助变量与被测量变量的关联度排序，实现变量精选。
- 需要指出，受系统自由度的限制，辅助变量的个数不能小于被估计变量的个数。至于辅助变量的最优数量，尚无统一结论。



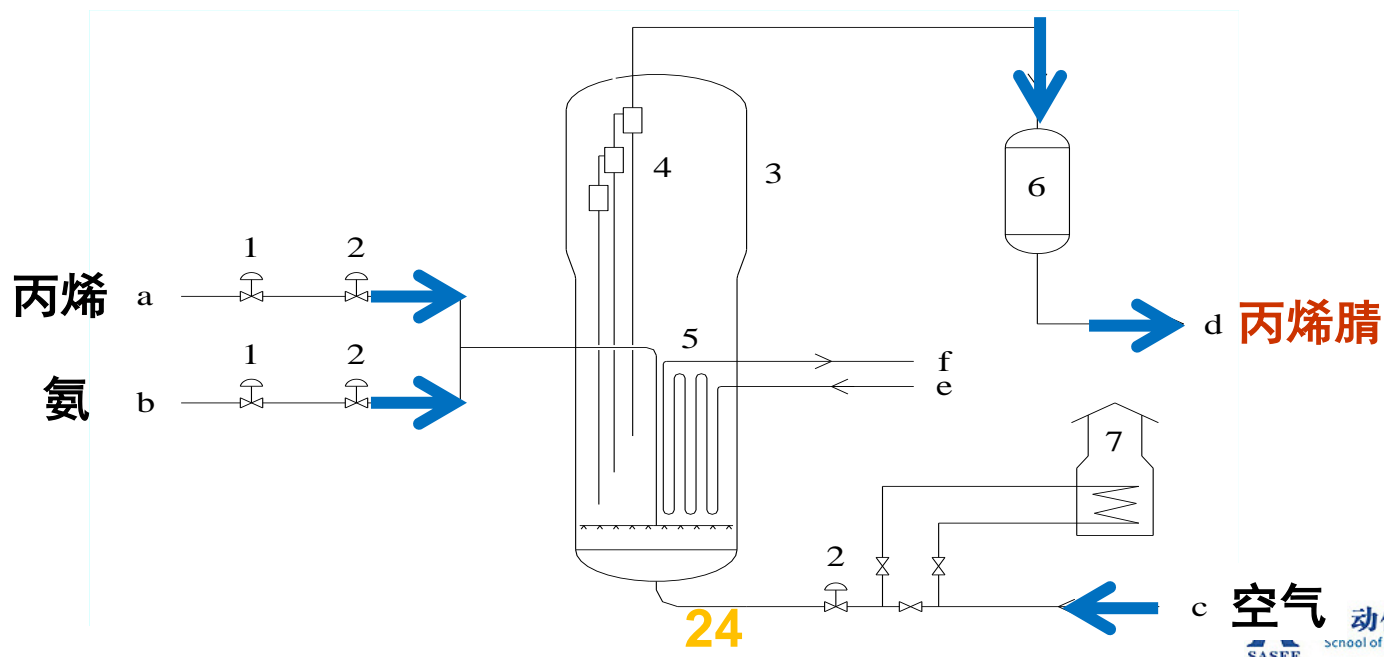
## 一种辅助变量的选择原则如下：

- **灵敏性：**能对过程输出(或不可测扰动)作出快速反应
- **特异性：**能对过程输出(或不可测扰动)之外的干扰不敏感
- **工程适应性：**工程上易于获得并达到一定的测量精度
- **精确性：**构成的估计器达到要求的精度
- **鲁棒性：**构成的估计器对模型误差不敏感

## 【例1】丙烯腈收率的在线测量

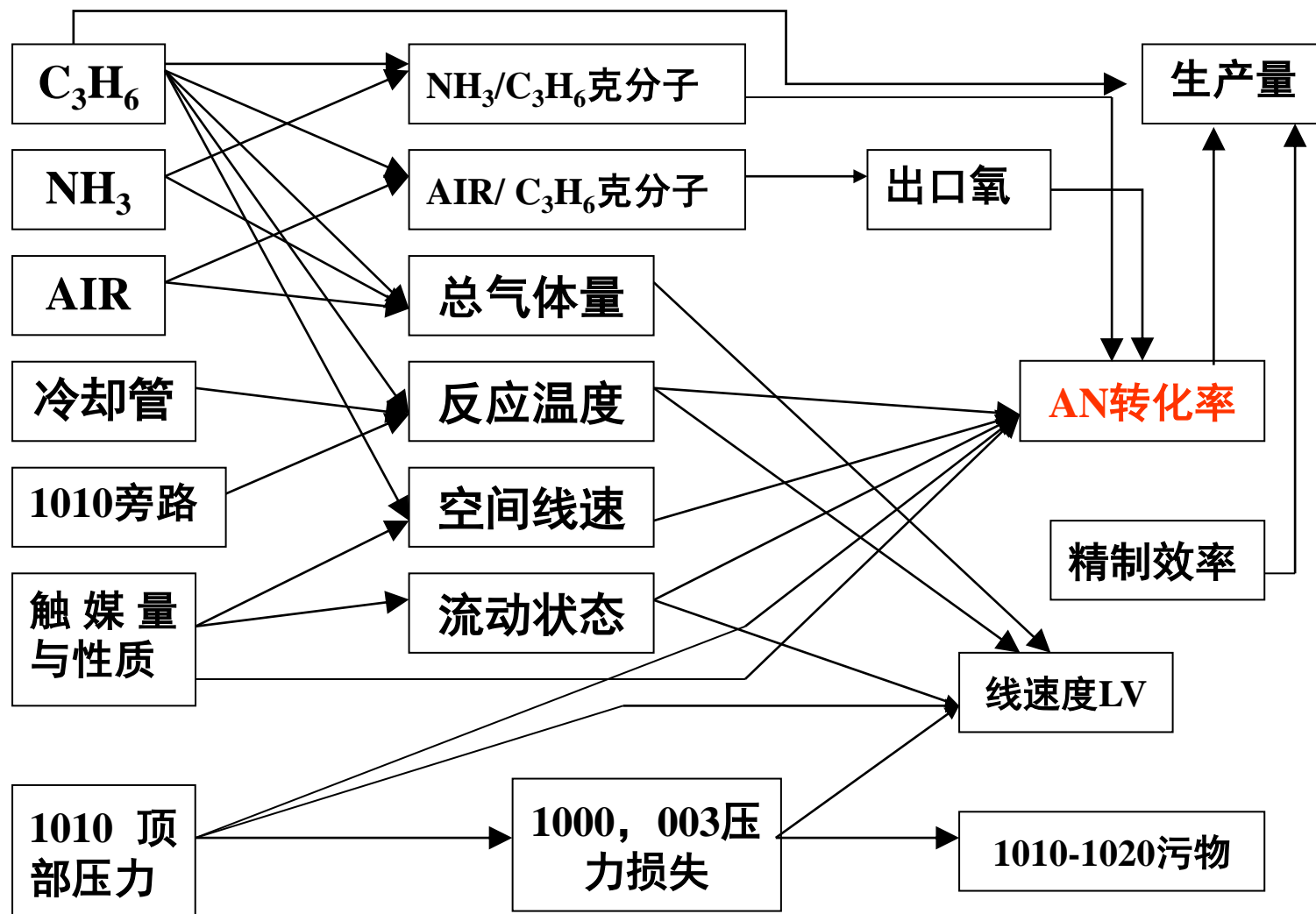
- 丙烯腈是丙烯系列的重要产品，既是生产有机高分子聚合物的重要单体，也是重要的有机合成原料。
- 通常采用丙烯氨氧化法生产丙烯腈，其主要原料为：
  - 丙烯
  - 氨
  - 空气

- 液态丙烯和液态氨经气化、过热、混合，由丙烯—氨分布器进入流化床反应器；
- 压缩空气通过位于丙烯—氨分布器下面的空气分布板进入流化床反应器；
- 在一定的温度和压力下，混合气体以一定的速度通过处于沸腾状的微球形催化剂床层；
- 在催化剂的作用下，**反应生成丙烯腈**。





- 流化床反应器是丙烯腈装置的心脏，其收率的高低直接影响整个装置的经济效益。
- 在工艺分析基础上，从反应器过程现场测量和分析参数中选择与丙烯腈收率密切相关的参数，建立基于几种不同方法的**丙烯腈收率软测量模型**，通过实验比较，确定其中软测量效果最好的一种实时测量模型投入日常运行，实现对丙烯腈收率的长期稳定的在线测量。
- 用软测量模型得到的丙烯腈收率精度与人工分析的精度接近，误差小于1%。





- 从现场可以采集到反应压力、中段温度、纯丙烯、空气比、氨比、触媒密度、触媒高度、触媒量、反应线速、尾氧浓度、丙烯腈收率及副产品氢氰酸及乙腈收率、未反应丙烯、一氧化碳收率、二氧化碳收率和未反应氨等**22个变量的数据**。
- 根据对丙烯腈工艺分析，将反应压力、中段温度、纯丙烯量、空比、氨比、反应线速、触媒量等7个变量选作为软测量建模的**辅助变量**，把丙烯腈收率作为**主导变量**。



## 【例2】采用主元分析法对辅助变量的选取分析：

- 现代工业过程往往需要测量很多过程变量，用以对过程进行检测和控制。而同一过程中的不同变量间往往相互关联。如在精馏塔中，进料组分的变化可以引起各塔板温度、塔顶和塔底组分等多个变量的变化。
- 如果能将很多相关的过程变量压缩为少数独立变量，就有可能从少数几个独立变量的变化中，较容易地找出引起过程变量错综复杂地变化的真正原因。
- **主元分析**：一种将多个相关的变量转换为少数几个相互独立变量的有效方法。

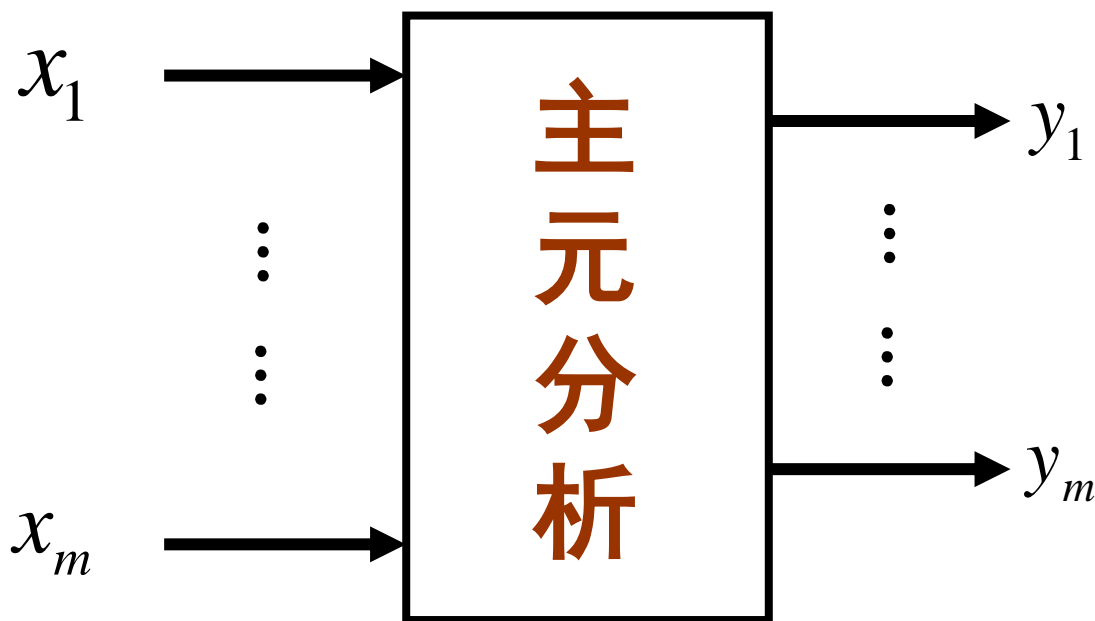


# 主元分析（PCA）

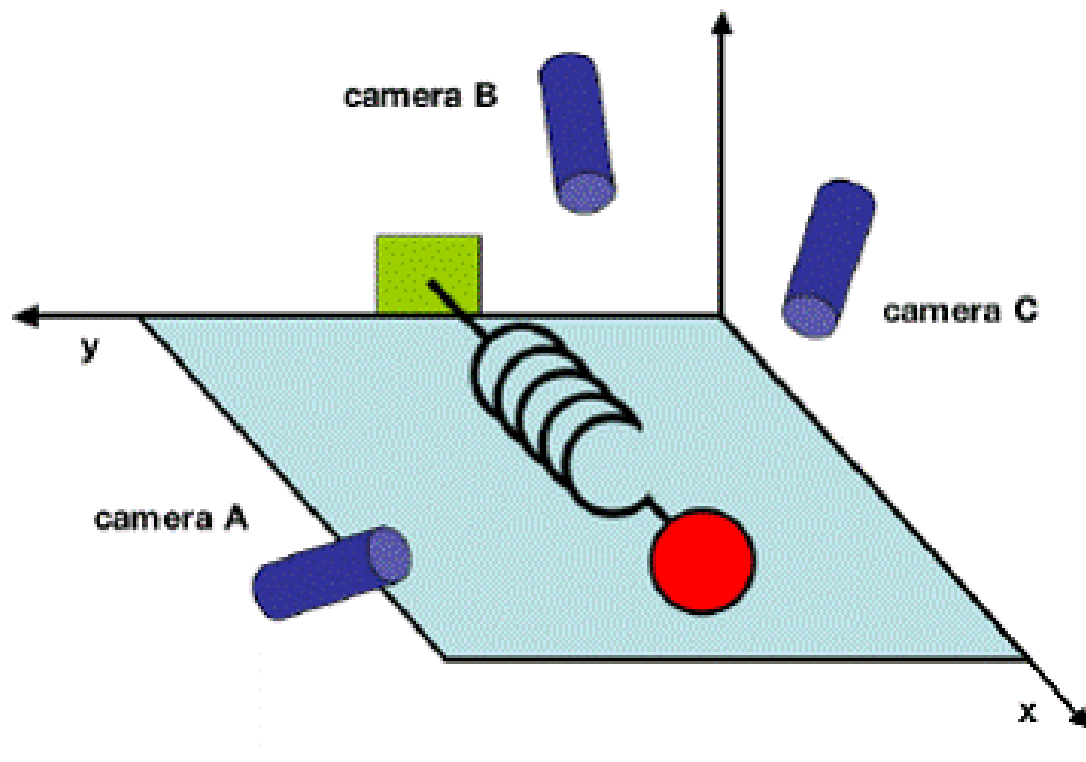
- 主元分析是由Pearson（1901）最早提出来的，用于研究对空间中的一些点进行最佳拟合直线和平面。
- 主元分析在化工过程中的应用始于80年代末和90年代初。
- 主元分析可以用来实现：数据简化、数据压缩、建模、奇异值检测、变量选择、分类和预报。

- 主元分析的基本算法：

- ① 找到一种空间变换方式，让（经标准化后的）原始变量**线性组合**成若干个向量，要求它们之间相互正交，且第一个向量能反映样本间自变量的最大差异，其他向量所反映的这种差异程度依次降低。
- ② 这些向量即为主元（Principal component）。



# 主元分析（PCA）的形象说明





# 主元分析 (PCA)

- 假设 $X$ 是一个 $n \times m$ 的数据矩阵（每一行对应一个样本；每一列对应一个变量），则矩阵可以分解为 $m$ 个向量的外积之和，即：

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_m p_m^T$$

$t_i \in R^n$  —— 得分向量( $X$ 的主元)  
 $p_i \in R^m$  —— 负载向量

- 写成矩阵形式为：

$$X = TP^T$$

$$T = [t_1, t_2, \cdots, t_n]$$

$$P = [p_1, p_2, \cdots, p_m]$$



# 主元分析 (PCA)

$$X = TP^T \quad \begin{array}{l} t_i \in R^n \text{ — 得分向量}(X\text{的主元}) \\ p_i \in R^m \text{ — 负载向量} \end{array}$$

- 各个得分向量之间是正交的，即：

$$\text{当 } i \neq j \text{ 时，满足 } t_i^T t_j = 0$$

- 各个负载向量之间也是正交的，且每个负载向量的长度都为1，即：

$$p_i^T p_j = 0 \quad i \neq j$$

$$p_i^T p_j = 1 \quad i = j$$

# 主元分析 (PCA)

- 对公式  $X = TP^T$  两侧同时右乘  $p_i$  , 可以得到:

$$Xp_i = t_1 p_1^T p_i + t_2 p_2^T p_i + \cdots + t_i p_i^T p_i + \cdots + t_m p_m^T p_i$$

- 可以得到:  $t_i = Xp_i$   $[i = 1, 2, \cdots m]$

得分  
向量

负载  
向量

$$t_i = Xp_i$$

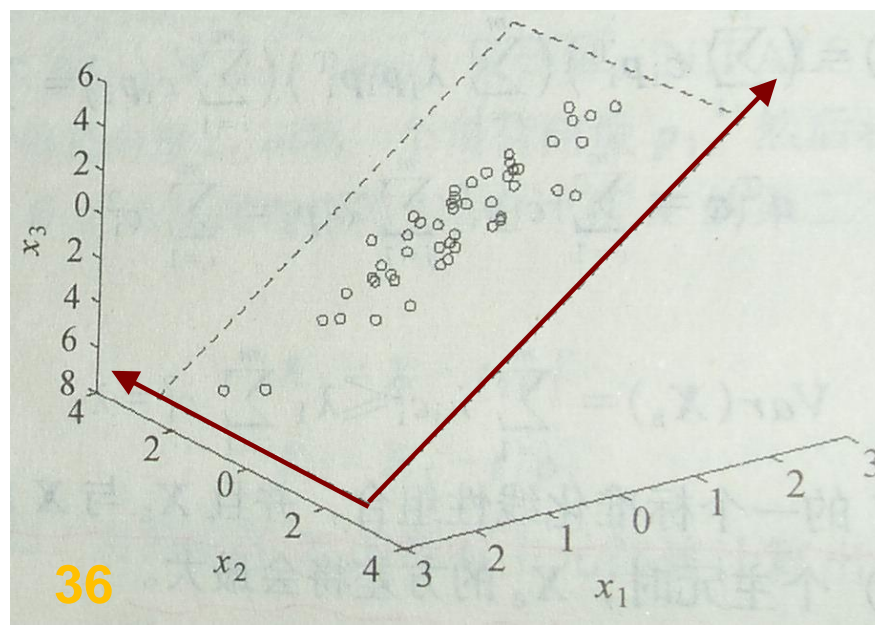
- 公式说明每一个得分向量实际上是数据矩阵X在和这个得分向量相对应的负载向量方向上的**投影**。
- 如果将得分向量按其长度做以下排列：

$$\|t_1\| > \|t_2\| > \cdots > \|t_m\|$$

- 那么负载向量  $p_1$  将代表数据X变化最大的方向， $p_1$  与  $p_2$  垂直并代表数据X变化的第二大方向， $p_m$  将代表数据X变化最小的方向。

## 主元分析的几何解释（以三维空间为例）

- 如果用一个主元来表示这些数据，可以用一条有向直线，即通过这组数据中心的一条直线来表示，它的方向是这些数据变化最大的方向；
- 如果用两个主元来表示这些数据，可以用两条相互垂直的直线构成的平面来表示，其中一条直线是数据的第1个主元，代表数据最大变化方向；另外一条直线是数据的第2个主元，代表数据的第二大变化方向。





- 当矩阵X中的变量间存在一定程度的线性相关时，数据X的变换将主要体现在最前面的几个负载向量的方向上，因此在最后面的几个负载向量上的投影会很小（主要是由于测量噪声引起的），这样就可以将主元分解后的矩阵写成：

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E$$

- E — 误差矩阵，由噪声引起，忽略后不会引起数据中  
有用信息的明显损失，从而可将数据近似表示为：

$$X \approx t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T$$

$$k < m$$



- 对矩阵X进行主元分析实际上等效于对X的协方差进行特征向量分析。

- 定义X的协方差矩阵为：

$$\text{cov}(X) = \frac{X^T X}{m-1}$$

- 对协方差矩阵进行正交分解：

$$\text{cov}(X) = P_m^T D P_m$$

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$$

$\lambda_i$  —— 协方差矩阵的特征值

$$P_m = [p_1, p_2, \dots, p_m]$$

$P_i$  —— 协方差矩阵的特征向量

- 如果将m个特征值按降序排列，即： $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ ，  
则与特征值对应的特征向量， $p_1, p_2, \dots, p_m$  即为矩阵X  
的负载向量。

- 主元分析主要应用于对高维空间进行降维
- 在降维过程中，必然要损失部分信息
- 需要研究降到多少维，才能保证主要信息没有遗漏
- 引入主元贡献率来衡量主元中所包含的信息量

主元贡献率

定义：

$$\lambda_k \left( \sum_{i=1}^m \lambda_i \right)^{-1} \text{ —— 第 } k \text{ 个主元的方差贡献率}$$

## 累积贡献率

$$\left( \sum_{i=1}^k \lambda_i \right) \left( \sum_{j=1}^m \lambda_j \right)^{-1} \quad \text{——前}k\text{个主元的累积贡献率}$$

累积贡献率表示前 $k$ 个主元所解释的数据变化占全部数据变化的比例，一般选取 $>85\%$ ，或通过交叉校验来决定主元个数。



## 主元分析的程序：

$[PC, LATENT, EXPLAINED] = PCACOV(X);$

$k = 1, 2, \dots, m$

$pc1 = PC(:, 1:k);$

$X1 = X * pc1;$

X— 数据矩阵 ( $n \times m$ )

X1 —主元分析后的数据矩阵 ( $n \times k$ )

PC —特征矩阵 ( $m \times m$ )

LATENT —显示特征值

EXPLAINED —每个特征值占比，用百分比表示，每一个主成分所贡献的比例

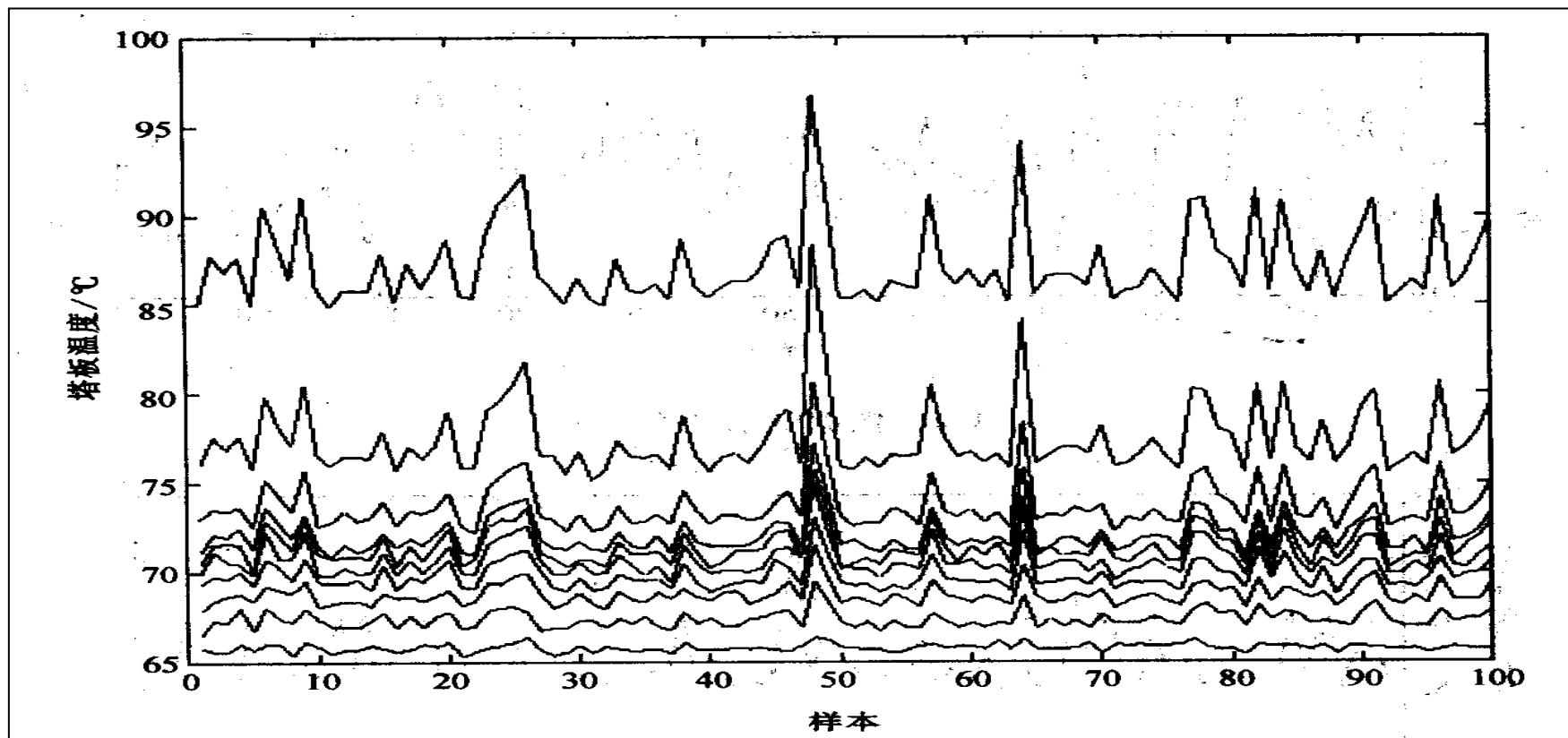


## 【例】丙烯腈收率数据的主元分析结果

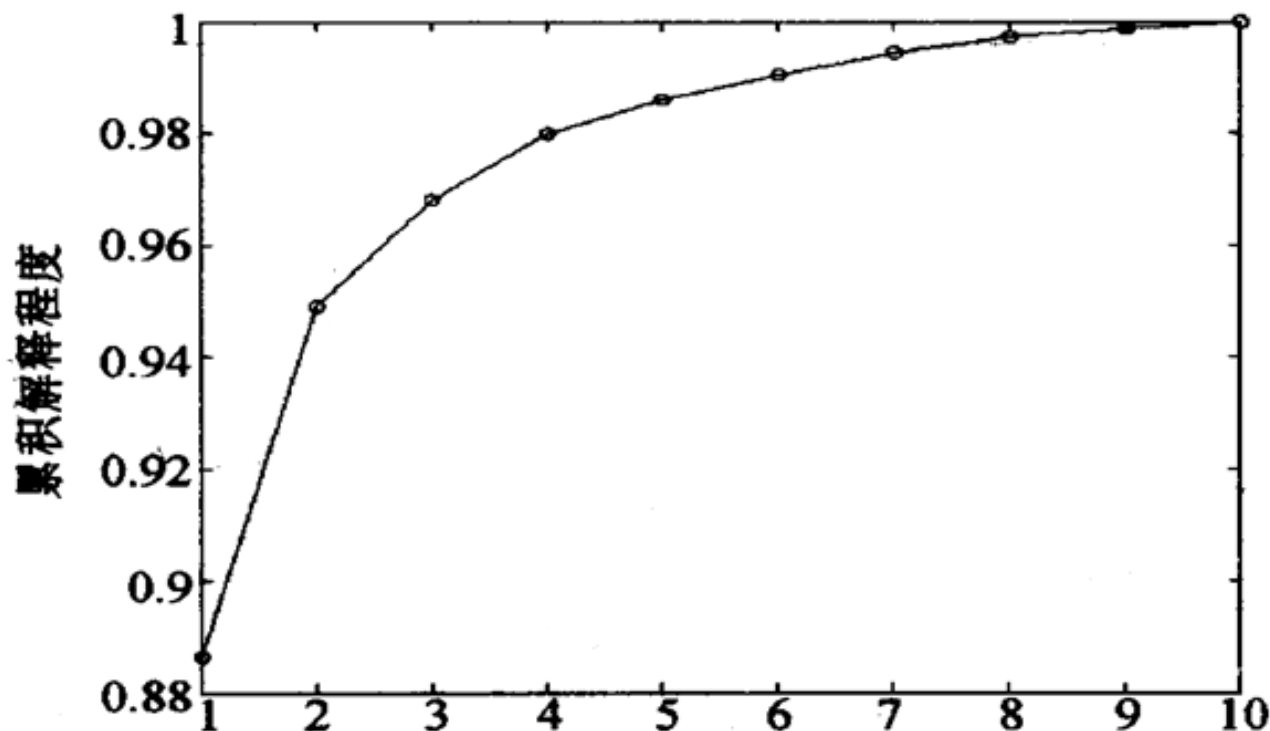
PCA模型的方差百分数			
主元数	特征值	方差/%	累积方差/%
1	36.539	26.04	26.04
2	31.189	22.227	48.267
3	20.421	14.553	62.82
4	19.369	13.803	76.623
5	18.461	13.157	89.78
6	12.386	8.8272	98.6072
7	1.9539	1.3925	100

# 主元分析举例

【例】从一个蒸馏塔的运行中采集10个塔板的100组温度。



- 对数据进行标准化处理，然后进行主元分析。
- 从图中可以看出：前两个主元能够解释超过95%的数据变化，因此可以将这10维数据压缩为2维。





## 5.2 测量数据的处理

软测量是依据易测过程变量与难以直接测量的待测过程变量之间建立数学关系，通过各种数学计算和估计方法，实现对待测过程变量的测量，测量数据的处理方法至关重要。

### 1) 误差处理

从现场采集的测量数据，受仪表精度和测量环境的影响，一般都不不可避免地带有误差，有时甚至有严重的过失误差。如果将这些现场测量数据直接用于软测量，会导致软测量精度降低，甚至完全失败。因此，测量数据必须经过误差处理。

**测量数据的误差：**随机误差、系统误差和粗大误差。



- 随机误差的处理

工程上多采用数字滤波算法。

如：算术平均滤波和一阶惯性滤波等。

随着计算机优化控制系统的使用，复杂的数值计算方法对数据的精确度提出了更高的要求，于是出现了数据一致性处理技术。

## • 粗大误差的处理

含有粗大误差的数据出现的机率较小，但是，一旦出现则可能严重破坏数据的统计特性，导致软测量的失败。

**提高测量数据质量的关键：**及时侦测、剔除和校正含有粗大误差的数据。

**侦测粗大误差的方法：**

- ①对各种可能导致粗大误差的因素进行理论分析
- ②借助于多种测量手段对同一变量进行测量，然后进行比较
- ③根据测量数据的统计特性进行检验等



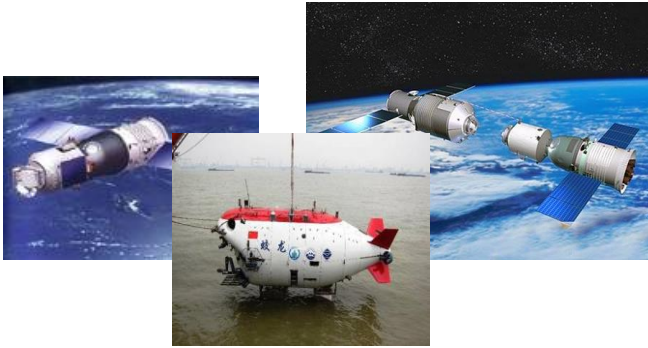
## 2) 数据的变换

对数据的变换包括**标度**、**转换**和**权函数**三方面：

- 工业过程中的测量数据有着不同的工程单位，直接使用这些数据进行计算，不能得到准确结果，甚至结果分散。利用合适的因子对数据进行**标度**，能够改善算法的精度和稳定性。
- **转换**包含对数据的直接转换以及寻找新的变量替换原变量两个含义。通过对数据的转换，可以有效地降低非线性特性。
- **权函数**则可实现对变量动态特性的补偿。合理使用权函数使我们有可能用稳态模型实现对过程的动态估计。







# Thanks

