

# 火电厂设备的健康状态监测、评价和诊断

## 摘要

随着我国工业现代化进程的加快，电力设备的重要性日益凸显，工业上对电力设备可靠性、安全性也已有了更高的要求。因此，对于设备运行状态的监测至关重要。本文通过研究火电厂设备的运行状态，研究各设备/设备群及各工序之间的关系，旨在通过客观的数据和科学的方法，对火电厂设备的健康状态进行监测、评价和判断。

首先，利用 MATLAB 软件，从海量数据中依据拉伊达准则进行数据剔除，提取健康样本，建立健康样本库。进行前后图像分析对比，得出健康样本较原始样本稳定性强，数据波动较小。其次，运用 MATLAB 曲线拟合工具箱对各参数间关联性和独立性进行分析，基于健康样本数据与理论知识，选择某一或某些参数作为主导因子，与其它参数进行拟合，建立参数回归方程，确定相关关系。结果表明汽泵前置泵电机 A 绕组温度与汽泵前置泵电流之间相关性最强，热膨胀和第一金属温度与汽轮机本体关联度较大，一次风机振动与风机电机电流相关性较强，这些因素有助于改善设备监测方案，降低数据处理维度。

建立健康样本库与参数间的关系后，进一步实现对当前运行状态的评估与对之后运行状态的预测。基于神经网络算法，建立数据学习与分析模型。该模型能够对健康库内数据进行自主学习，快速输出样本的正常值，实现单个设备的在线状态评估与诊断。最终结合原始数据，通过相关工具总结出异常情况的大致条件，并利用导数知识，准确发现各工况下设备的异常情况，及时给出报警信息。

**关键词：** 拉伊达准则 曲线拟合工具箱 参数回归方程 神经网络算法 自主学习

## 一、问题重述

目前电力系统的电厂均已建立实时监控信息系统，系统设备产生大量实时监控信息和设备正常运转时的历史数据。采用合理手段对生产运作的海量数据进行挖掘处理，建立生产设备的实时监控模型，实现设备、设备群或系统乃至机组的在线状态评估与诊断，同时找到改善系统、提高生产效率的方法，对于我国电力系统的发展进步具有重要意义。附件 1 给出火电厂设备智能诊断系统结构流程。附件 2—4 给出火电厂部分设备实施监测数据和设备正常运转时的历史数据。以某火电厂为例，试建立数学模型求解下列问题：

- 1、根据附件 2—4，建立从海量数据中筛选和提取设备健康样本库的算法。
- 2、基于健康样本数据，建立分析特征参数间关联性和独立性的算法，判断设备检测特征是否冗余，给出相应设备的优化检测特征。
- 3、根据附件 2—4，建立设备当前运行状态的动态综合评价模型，给出设备综合评价指标的算法以及设备在实时/历史工况下的正常值。
- 4、建立多设备群构成的系统模块、多模块构成的发电机组，以及拥有多机组的整个火电厂的实时监控模型，实现各层级的在线状态评估与诊断。

## 二、问题分析

### 2.1 问题 1

题目要求给出从海量数据中筛选和提取设备健康样本库的算法。通过查阅相关资料得知汽泵前置泵正常工作时，电流与功率是相对稳定的，因此以汽泵前置泵电流&实际值为主体进行操作；对于汽轮机本体，以发电机功率为主体进行操作；对于一次风机，以 1 号一次风机电机电流进行操作。基于本题附件中提供了大量数据，采用拉依达准则利用 MATLAB 软件剔除异常数据。

### 2.2 问题 2

基于问题一的健康样本数据，根据汽泵前置泵、汽轮机和一次风机的工作原理，分别选取汽泵前置泵电流、汽轮机发电功率以及 1 号一次风机电机电流作为主导因子（因变量），利用 MATLAB 的曲线拟合工具箱，对附件 2—4 中的因变量

与其他因子进行相关性分析,根据分析结果判定冗余的检测数据,优化监测特征。判定依据如下。

- 和方差 (SSE), 计算公式为:

$$SSE = \sum_{i=1}^n \omega_i (y_i - \hat{y}_i)^2 \quad (2.1)$$

结果越接近于 0, 模型选择和拟合更好。

- 决定系数 (R-square), 计算公式为:

$$R - \text{square} = SSR/SST = (SST - SSE)/SST = 1 - SSE/SST$$

通过数据的变化来表征一个拟合的好坏。决定系数的正常取值范围为 [0, 1], 越接近 1, 表明方程的变量对 y 的解释能力越强, 这个模型对数据拟合的也较好。

- 均方根误差 (RMSE), 也称回归系统的拟合标准差, 计算公式:

$$RMSE = \sqrt{MSE} = \sqrt{(SSE/n)} = \sqrt{(\sum_{i=1}^n \omega_i (y_i - \hat{y}_i)^2/n)} \quad (2.2)$$

RMSE 越小, 拟合效果越好。

### 2.3 问题 3

问题三需要根据检测特征的实时运行数据和设备正常运转时的历史数据, 建立当前运行状态的动态综合评价模型。对该问题的求解可以分为数据学习和数据分析两部分。

数据学习部分。在这部分中, 要求算法能够对健康数据库内的数据进行学习, 找出各观测值与健康状况指标之间的关系。为了达到这个目的, 可以使用神经网络算法, 将问题一中筛选出来的数据分为两部分, 一部分用作神经网络的训练, 另一部分用作训练效果检测, 通过多次训练, 得到拟合情况比较好的神经网络。

数据分析部分。在这个部分中, 要求对实时得到的数据进行分析, 评估设备当前的状态。通过问题 1 中对大量数据的分析, 我们可以采取取平均值的方法的得到每个设备的标准健康数值。当设备健康值与标准值不同的时候, 我们将根据两者之间的差值大小划定不同的范围, 确定设备的运行状况。将实时数据传入已经训练好的神经网络, 得出设备当前的健康值, 通过将当前健康数值与标准健康数值的比较与分析我们就可以确定设备当前所处的状态。

### 2.4 问题 4

问题四需要在单个设备的评估与诊断的基础上,根据相关资料建立系统模块、发电机组乃至整个火力发电厂的监控模型。针对不同设备之间的函数关系和比例系数的确定,我们可以通过查阅相关资料来进行定性的分析。该模型主要有报警和状态分析查询功能。

报警功能的实现主要基于当前状态对未来状态的预测,由于设备健康值与时间并无直接的函数关系,基于假设(设备健康值是随时间渐变的,不存在突变情况),所以在短时间内将健康值看作以一定速率随时间渐变,从而可以实现对健康状态的短期预测。使用问题3中得到的分析模型对短期预测值进行分析,决定是否报警。

状态分析查询功能,我们通过选取某一时间段的检测数据,分析其中各种状态所占的比例,通过给每个状态赋值的方法,可以计算得出一个评价价值,从而对该时间段内设备的运行状态进行定量的分析。

### 三、符号说明

AQ	气泵前置泵电流标准误差
BQ	汽轮机发电功率标准误差
CQ	一次风机电流标准误差
Aav	汽泵前置泵电流平均值
Bav	汽轮机发电功率平均值
Cav	一次风机电机电流平均值
&	相似度
¥	异常度
V	剩余误差

### 四、模型假设

1. 数据来源真实准确
2. 各变量之间相互独立,互不影响
3. 设备状态随时间的变化均匀,不存在状态突变的情况
4. 处于不同系统的设备之间不存在相互影响
5. 数据中设备正常运行的样本居多
6. 除了数据中的特征参数外,忽略其他因素对设备健康值的影响

### 五、建模过程

## 5.1 问题 1 的建模和求解

### 5.1.1 选择衡量标准

#### 1. 汽泵前置泵

通过查阅汽泵前置泵的相关资料，得知随着机组负荷的升高，给水流量增加，前置泵电机电流上升。如前置泵运行中发生两台电机电流偏差较大和单台电机过流、过热现象，长时间运行，将会严重危及设备安全，制约机组带负荷能力。因此，我们认为可以理想地将前置泵电流作为衡量前置泵正常工作的标准。

#### 2. 汽轮机本体

由相关资料可知，汽轮机的主要作用为将锅炉生产蒸汽热能转化为转子旋转机械能。从其最终目的出发，我们理想化的将其发电机功率作为衡量汽轮机本体正常工作的标准。

#### 3. 一次风机

查阅资料得知，一次风机的主要作用为提供一定压力、一定流量的一次风，将煤粉干燥并送入喷燃器，提供煤粉挥发份燃烧所需热量。而风机电流和风量成正比，就是加大风量，就要加大风机的动力即电流。因此，我们理想化的将其电流作为衡量一次风机正常工作的标准。

### 5.1.2 数据的过滤

#### 1. 方法的选择

题目附件所给数据量极大，因此我们联想到可以利用拉依达准则来剔除异常数据。拉依达准则是指先假设一组检测数据只含有随机误差，对其进行计算处理得到标准偏差，按一定概率确定一个区间，认为凡超过这个区间的误差，就不属于随机误差而是粗大误差，含有该误差的数据应予以剔除。这种判别处理原理及方法仅局限于对正态或近似正态分布的样本数据处理，它是以测量次数充分大为前提的，当测量次数小的情形用准则剔除粗大误差是不够可靠的。

### 5.1.3 拉依达准则

由于三种设备利用拉依达准则的方法大同小异，仅是数据不同而已。下面以汽泵前置泵为例，详细介绍如何剔除异常数据。将实时监测到的前置泵电流数据，输入 matlab 中。通过计算得到下图：

A	3940x7 double
A2	3203x7 double
Aav	43.4447
Anet	1x1 network
APoutput	1203x1 double
AQ	5.3894
Axiangsi	3940x1 double
Ayuce	3940x1 double
Azhuangtai	3940x1 double

图 1.1

再依次将表中数据与 AQ 进行比较，若剩余误差  $|v|=|X_n - A_{av}| > AQ$ ，则认为  $X_n$  是含有粗大误差值的坏值，应予剔除。并同时将其同一时间得到的汽泵前置泵电机 A 绕组温度 2 与实际值等六组数据同时剔除。经此剔除操作后，所剩数据即可作为汽泵前置泵的健康样本库。

如下是数据剔除前后曲线图：

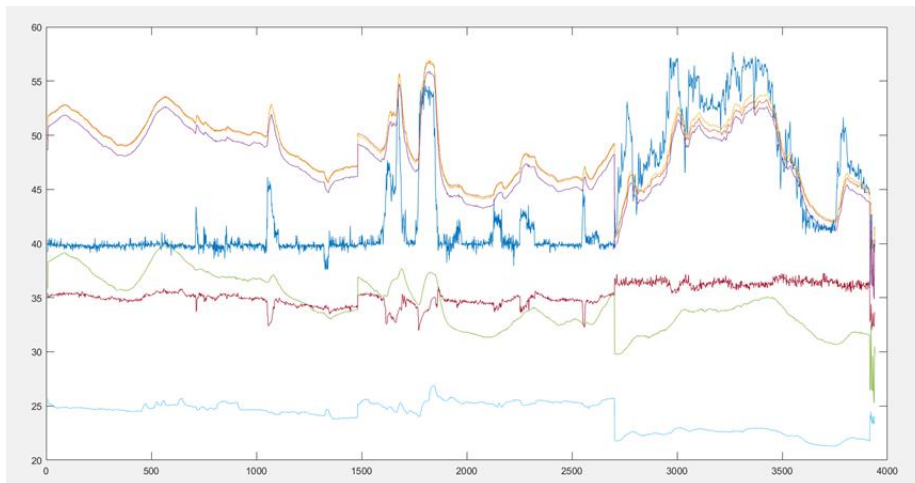


图 1.2 原始样本曲线图

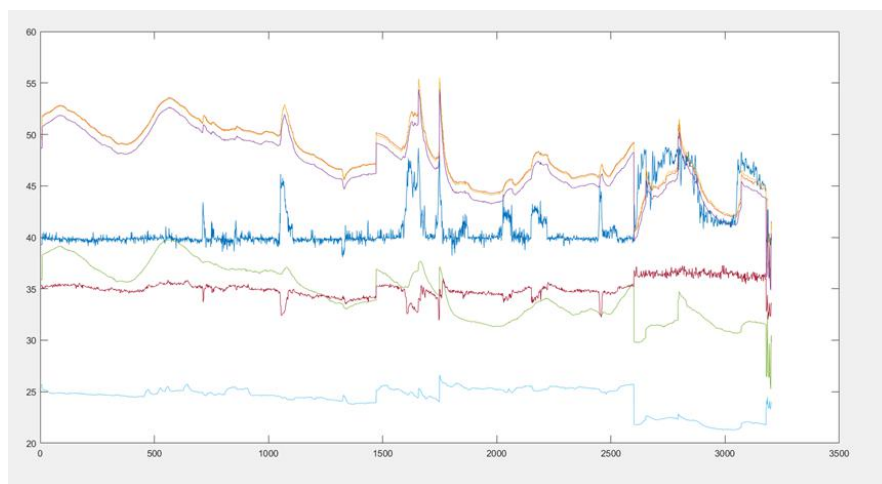


图 1.3 健康样本曲线图

明显可见，经过拉依达准则剔除后，曲线峰值与波动小了很多。达到预期效果。

## 5.2 问题 2 的建模与求解

## 5.2.1 汽泵前置泵监测数据分析

### 1. 主导因子选取分析

由于给水泵入口的水温接近该压力下的饱和温度，工作条件恶劣，很容易发生汽化，为保证给水泵的安全运行，为给水泵配置流量与给水泵相匹配的前置泵。因前置泵转速低，抗汽蚀性能好，除氧器的给水经前置泵增压后进入给水泵，保证了锅炉给水泵所需的足够的汽蚀余量，从而大大改善了给水泵抗汽蚀的性能。

但随着机组负荷的升高，给水流量增加，前置泵电机电流上升。如前置泵运行中发生两台电机电流偏差较大和单台电机过流、过热现象，长时间运行，将会严重危及设备安全，制约机组带负荷能力。

因此选择汽泵前置泵电流作为因变量  $y$ ，汽泵前置泵电机 A 绕组温度、电机 B 绕组温度、电机 C 绕组温度、汽泵前置泵电机径向轴承温度、汽泵前置泵自由端径向轴承温度、汽泵前置泵传动端径向轴承温度，作为自变量  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$ 、 $x_5$ 、 $x_6$  分别进行拟合，生成下列图表和数据。

### 2. 图表分析

· 模型选取：

$$\text{Linear Fitting } y = a * \sin(x - \pi) + b * (x - 10)^2 + c \quad (5.1)$$

$$\text{Fourier } y = a_0 + a_1 * \cos(x\omega) + b_1 * \sin(x\omega) \quad (5.2)$$

表 1.1

自变量对应表	
符号	名称
x1	汽泵前置泵电机A绕组温度
x2	汽泵前置泵电机B绕组温度
x3	汽泵前置泵电机C绕组温度
x4	汽泵前置泵电机径向轴承温度
x5	汽泵前置泵自由端径向轴承温度
x6	汽泵前置泵传动端径向轴承温度

表 1.2

汽泵前置泵监测数据分析										
自变量xi	回归系数							和方差	决定系数	均方根误差
	a	b	c	a0	a1	b1	w			
x1	-38.39	1.515	-154.6					174.4	0.6273	0.2105
x2	-33.52	1.339	-136.3					248.1	0.4698	0.251
x3	-45.46	1.894	-191.9					193.8	0.5857	0.2219
x4				-390900	390900	-769.1	0.004024	253.3	0.4586	0.2537
x5	-6.71E+24	-4.67E+22	1.13E+25					1538	-2.288	0.6251
x6				0.7107	0.0332	0.3839	6.568	268.9	0.4254	0.2614

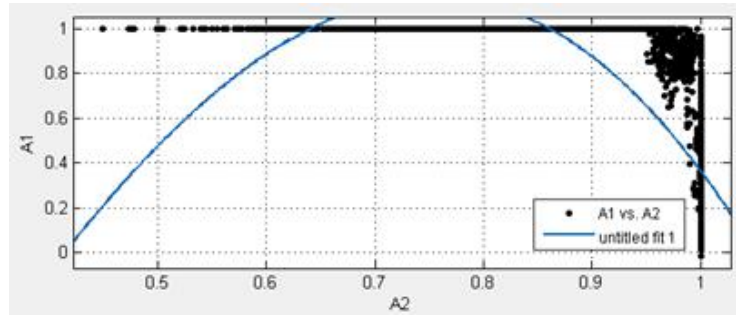


图 1.4 汽泵前置泵电流与电机 A 绕组温度

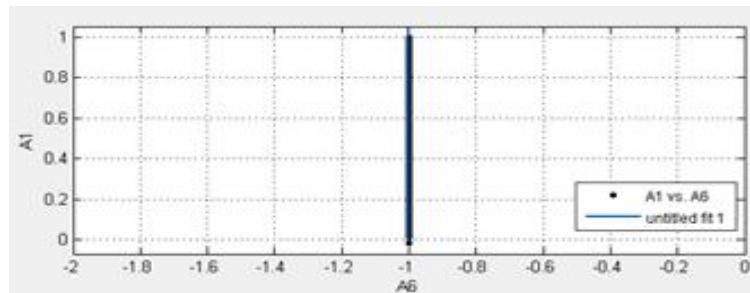


图 1.5 汽泵前置泵电流与自由端径向轴承温度

其中 $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_5$ 采取 Linear Fitting 模型， $x_4$ 、 $x_6$ 采取 Fourier 模型。由表中数据可知， $x_1$ 与  $y$  的决定系数最大，方差最小，拟合效果最好，关联度也最大；反之， $x_5$ 与  $y$  的决定系数 $<0$ ，方差 $>1000$ ，拟合程度最差，关联度最小。其余自变量与因变量的相关度依次是 $x_3 > x_2 > x_4 > x_6$ 。因此，自变量 $x_5$ 即汽泵前置泵自由端径向轴承温度有部分冗余特征，可做弱化或削减处理。

## 5.2.2 汽轮机本体监测数据分析

### 1. 主导因子选择分析

汽轮机是利用蒸汽的热能进行做功的旋转机械，热能转换为机械能推动汽轮机做功进而发电，因此选取汽轮机发电功率作为因变量  $y$ ，其余因素作为自变量  $x_i$  ( $i=1, 2, 3, \dots, 22, 23$ ) 分别进行拟合，生成下列图表和数据

### 2. 图表分析

· 模型选取：

$$\text{Fourier} \quad y = a_0 + a_1 * \cos(x\omega) + b_1 * \sin(x\omega) \quad (5.3)$$

$$\text{Linear Fitting} \quad y = a * \sin(x - \pi) + b * (x - 10)^2 + c \quad (5.4)$$

$$\text{Polynomial} \quad y = p_1 * x + p_2 \quad (5.5)$$

表 1.3



自变量对应表	
符号	名称
x1	发电机端中压上下壁温差
x2	调速器端中压上下壁温差
x3	高压上下壁温差
x4	1号送风机入口风温
x5	2号送风机入口风温
x6	2#高压主汽门前蒸汽温度(右)
x7	1#高压主汽门前蒸汽温度(左)
x8	再热蒸汽温度
x9	热膨胀1
x10	热膨胀2
x11	发电机端高中压缸端壁温度
x12	调速器端高中压缸端壁温度
x13	第一级金属温度
x14	汽室内左壁温
x15	汽室外左壁温
x16	汽室内右壁温
x17	汽室外右壁温
x18	发电机端中压下缸检水
x19	调速器端中压下缸检水
x20	发电机端中压上缸检水
x21	调速器端中压上缸检水
x22	高压下缸检水
x23	高压上缸检水

表 1.4

自变量xi	汽轮机本体监测数据分析											和方差	决定系数	均方根误差
	回归系数													
	a0	a1	b1	w	a	b	c	p1	p2					
x1								4.247e+10	4.247e+10			149.30	-9.577e-06	0.1523
x2	0.2139	0.3895	-0.5059	41.58								41.26	0.7235	0.0801
x3	-0.1639	0.1449	-0.1761	121.30								124.40	0.1666	0.1391
x4	-0.2943	-0.0375	-0.0629	114.30								138.60	0.0711	0.1468
x5	-0.3202	-0.0297	-0.0377	139.40								143.90	0.0356	0.1496
x6	-0.3353	-0.0114	0.0239	248.70								146.80	0.0163	0.1511
x7	-0.2929	-0.0681	-0.0268	115.10								139.20	0.0675	0.1471
x8	-0.2978	-0.0142	-0.0777	88.43								142.10	0.0479	0.1487
x9	-0.3639	-0.0811	0.0444	303.60								130.70	0.1246	0.1425
x10	-0.1522	-0.0398	0.3119	176.70								39.31	0.7366	0.0782
x11	-0.1934	-0.0744	-0.2049	58.72								56.56	0.6210	0.0938
x12					1018	-51.09	5108					140.40	0.0596	0.1477
x13	-0.1722	-0.1423	0.2930	17.87								36.66	0.7544	0.0755
x14	-0.2752	-0.0470	-0.0907	60.69								141.30	0.0532	0.1482
x15	-0.2610	-0.0992	0.1074	73.08								126.20	0.1548	0.1401
x16	-0.2573	0.1005	0.1252	71.58								111.10	0.2557	0.1314
x17	-0.2590	-0.0852	-0.1539	68.91								69.51	0.5343	0.1040
x18	-0.2128	0.0922	-0.1926	67.51								57.26	0.6163	0.0944
x19	-0.2822	-0.0837	-0.0074	80.65								139.10	0.0680	0.1471
x20	-0.2443	0.0370	0.1607	85.31								88.66	0.4060	0.1174
x21	-0.2341	0.0465	0.1846	77.80								83.68	0.4394	0.1141
x22	-0.1917	0.3250	-0.0317	25.75								38.31	0.7433	0.0772
x23	-0.1908	0.2780	-0.1438	25.20								32.77	0.7804	0.0714

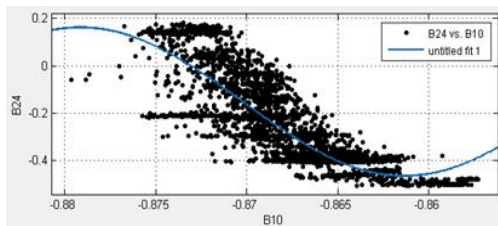


图 1.6 汽轮机发电功率与热膨胀 2

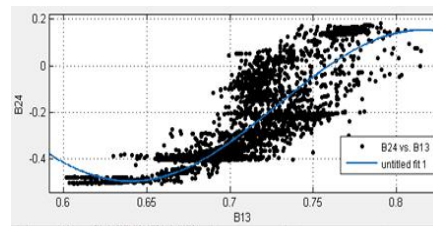


图 1.7 汽轮机发电功率与第一级金属温度

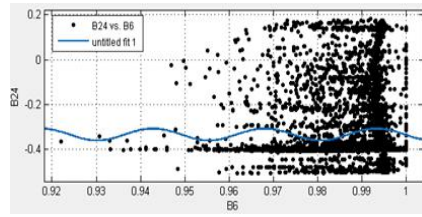
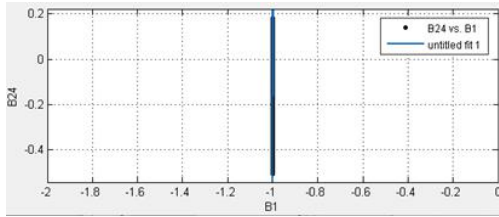


图 1.8 汽轮机发电功率与发电机端中压上下壁温差 图 1.9 汽轮机发电功率与 2#高压主汽门前蒸汽温度(右)

其中 $x_1$ 采用 Polynomial 模型， $x_{12}$ 采用 Linear Fitting 模型，其余变量均采用 Fourier 模型。通过拟合可知， $x_{10}$ 与  $y$  的和方差为 39.31，决定系数为 0.7366， $x_{13}$ 与  $y$  的拟合方差为 36.66，决定系数为 0.7544，拟合程度较好，相关性较大； $x_1$ 、 $x_6$ 与  $y$  的拟合程度较差，方差分别为 149.30、146.80，决定系数分别是 $-9.577e-06$ 、0.0163，均方根误差为 0.1523、0.1511，关联程度较弱。因此， $x_1$ 与 $x_6$ 即发电机端中压上下壁温差和 2#高压主汽门前蒸汽温度(右)可作为冗余监测数据处理。

### 5.2.3 一次风机监测数据分析

#### 1. 主导因子选择分析

发电厂一次风机的作用是给提供一定压头和流量的空气，携带制粉系统出口的煤粉进入炉膛进行燃烧，同时也能与二次风协调作用为煤粉燃烧提供助燃。通过监测风机电机电流的变化实时监控一次风机的运转情况，及时调整风机功率与风力大小，便于改变煤粉的鼓入量和燃烧程度。因此，选取 1 号一次风机电机电流作为因变量  $y$ ，风机振动 1、振动 2、轴温 1、轴温 2 分别作为自变量，生成以下图表信息。

#### 2. 图表分析

· 模型选取：

$$\text{Fourier} \quad y = a_0 + a_1 * \cos(x\omega) + b_1 * \sin(x\omega) \quad (5.6)$$

$$\text{Linear Fitting} \quad y = a_2 * \sin(x - \pi) + b_2 * (x - 10)^2 + c_1 \quad (5.7)$$

表 1.5

自变量对应表	
符号	名称
x1	1号一次风机振动1
x2	1号一次风机振动2
x3	1号一次风机轴温1
x4	1号一次风机轴温2

表 1.6

1号一次风机监测数据分析													
自变量	回归系数										和方差	决定系数	均方根误差
	a0	a1	b1	w	a2	b2	c1	a3	b3	c2			
x1								-161.00	5.03	-472.20	238.30	0.7989	0.2192
x2								-169.50	5.17	-482.30	249.60	0.7894	0.2244
x3					-15.98	0.71	-70.85				1093.00	0.0775	0.4695
x4	0.80	0.44	-0.12	9.82							1087.00	0.0827	0.4683

自变量 $x_1$ 、 $x_2$ 、 $x_3$ 均采用 Linear Fitting 模型， $x_4$ 选取 Fourier 模型。读表可知， $x_3$ 、 $x_4$ 的和方差均大于 1000，决定系数小于 0.1，拟合效果极差，关联度极小，因此可以考虑取消监测 1 号一次风机轴温 1、轴温 2 两组数据，提高监测数据精准度，减少冗余。

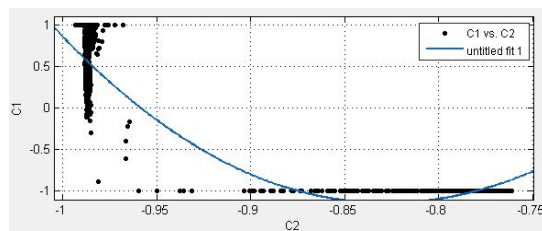


图 2.0 1 号一次风机电机电流与振动 1

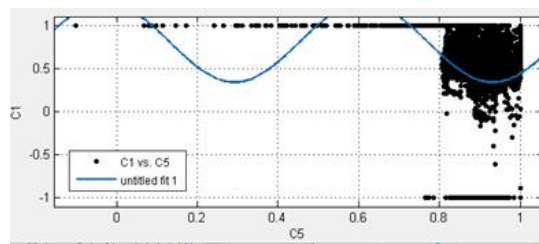


图 2.1 1 号一次风机电机电流与轴温 2

### 5.3 问题 3 的建模与求解

通过第一问的求解，剔除了三组数据中的异常数据，得到了健康数据库 A2、B2、C2，数据库内的数据反映了设备正常运行情况，可以用来进行神经网络的训练。

当神经网络训练完成之后可以使用它来分析检测数据，得到设备运行健康值，进行运行状态分析。

#### 1. 神经网络训练与训练效果测试：

##### (1) 训练：

由于神经网络训练需要用到大量的现有数据，所以将问题一中所得到的健康数据样本的大部分用作训练样本。其中矩阵 A2 的前 2000 项用作训练样本，矩阵 B2 的前 4000 项用作训练样本，矩阵 C2 的前 3000 项用作训练样本。

由于目前的研究对于神经网络算法中神经元数目的确定没有有效的方法，只

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/768031135020006045>