

---

(2017 泰迪杯全国大学生数据挖掘竞赛 B 题)

# 关于中央空调系统的数据分析与控制策略

——太和顾问杯数学建模决赛论文

姓名：高 齐

学号：14061128

学院：数理学院

原创声明：本论文所有内容均为原创，未经允许禁止转载。

—— 2017 年 5 月 1 日

---

## 摘要

本文通过建立数学模型的方法对中央空调系统的数据进行非线性回归，再结合给定的条件进行最优化分析和设计。

对于问题一基于曲线拟合的原理，使用 **1stopt** 软件对实际数据做深入的分析处理，挖掘出数据变化的特征和规律。本模型使用拟牛顿算法（**BFGS**），研究冷却负载、系统效率、耗电量与可控变量和不可控变量之间的关系模型，并检验模型的可靠性，即算出变量之间的函数表达式，并得出模型的误差分析：均方根误差，误差项平方和，相关系数，复相关系数，决定系数。之后进行分析和优化，找出最佳方案

关键词：曲线拟合，**1stopt**，拟牛顿算法（**BFGS**），误差分析，全局优化。

## 一， 问题重述

### 一：问题的背景

随着全球气候的变迁和空调技术的发展，越来越多的大型建筑物利用中央空的快速推进，如何围绕智慧城市建设实现中央空调系统的智能控制与节能，这是智慧城市建设中的重要研究课题之一。中央空调系统的优化控制策略研究也是实际中的一个很有普遍意义的重要课题。图 1 给出了常见的一类中央空调系调系统来实现室内温度和湿度的调节控制。特别是随着“智慧城市”建设步伐统的基本结构示意图，该系统包括三套冷却装置 Chiller，记为 CH-1/2/3）、两个冷却塔（Cooling Tower，记为 CT-1/2，二者等效）、三个冷凝水泵（Condenser Water Pump，记为 CWP-1/2/3）和四个冷水泵（Chilled Water Pump，记为 CHWP1/2/3/4）。三套冷却装置的额定功率分别为 550RT，550RT 和 235RT（RT 为冷却吨，即表示制冷能力的功率单位，1 RT = 3.517kw）。

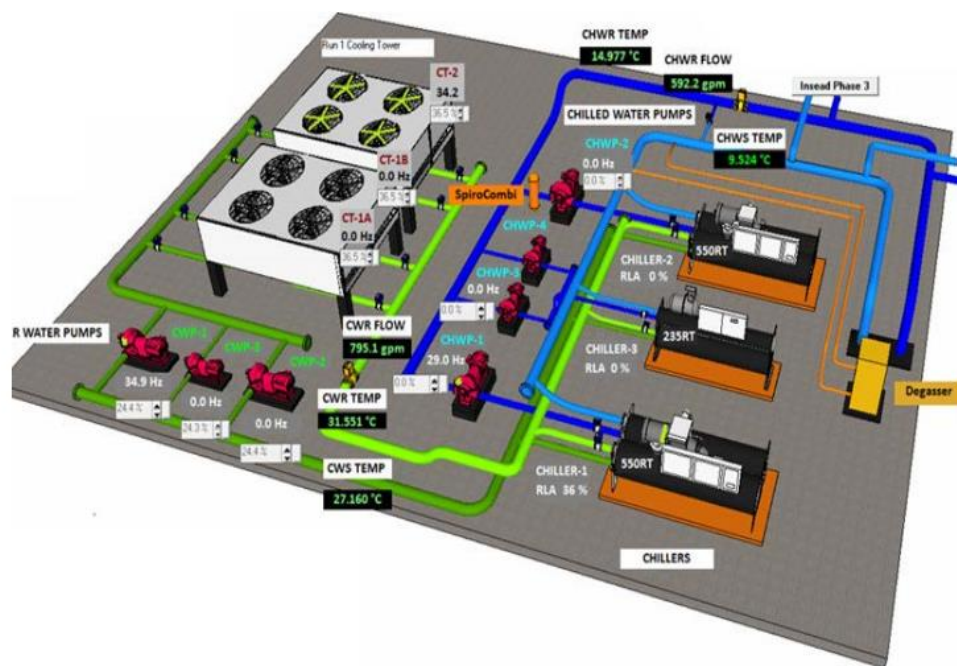


图 1. 中央空调系统的基本结构示意图

图 2 给出了中央空调系统的基本工作原理图。每一套(水冷)中央空调系统都包含内循环和外循环两个热交换循环系统。在内循环(图 2 下方)中, 冷水泵将冷却装置中由冷却器冷却的冷水推进大楼, 通过热交换对大楼内部的空气进行降温 and 除湿。循环水在吸收了室内空气中的热量以后温度升高, 重新回流至冷却器中冷却降温, 并通过冷却装置将其热量传送到外循环。在外循环(图 2 上方)中, 冷凝器水泵推动冷凝器中的水来吸收冷却器降温所产生的热量到冷却塔, 冷却塔把水中的热量排放到室外空气中, 水流再流回冷凝器。依次循环。内循环中的冷却器和外循环中的冷凝器被封装在一起, 称为中央空调系统的冷却装置(Chiller)。中央空调通过能量转换实现将室内的热量 2 吸收并输送至室外, 从而实现换气降温的功效。

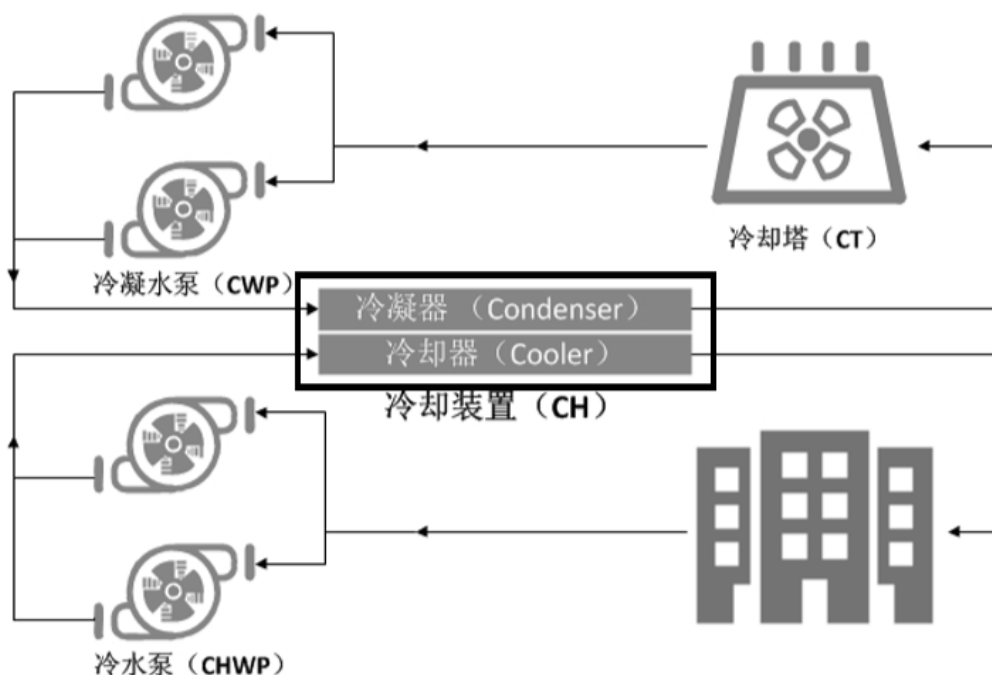


图 2. 中央空调系统的工作原理图。

## 二. 问题的数据说明

1.附件 1: CACS\_data.csv (共 51 个字段, 88840 条记录, 数据样本含 20593 条记录)给出了一个实际中央空调系统状态参数和传感器采集的数据。包括以下内容: (1) 采集时间: 年月日时分; (2) 环境信息: 室外相对湿度和温度; (3) 系统状态参数: 冷水泵状态、冷凝泵状态、冷却装置状态和冷却塔状态等; (4) 控制参数: 冷水泵转速、冷凝水泵转速、冷却塔风扇转速等, 相应转速的频率范围分别为  $30\text{Hz} \sim 50\text{Hz}$ 、 $30\text{Hz} \sim 50\text{Hz}$  和  $25\text{Hz} \sim 50\text{Hz}$ 。数据中将以百分比%的形式给出, 比如: 以  $30\text{Hz} \sim 50\text{Hz}$  为例,  $0\% = 30\text{Hz}$ ,  $100\% = 50\text{Hz}$ ,  $50\% = 40\text{Hz}$ 。(5) 系统采集信息: 设备的耗电量、相关传感器读数(如冷凝水进/出冷却装置的水温和流速等); (6) 系统运行相关信息: 耗电量、冷却负载、系统效率等。

2.附件 2: 关于数据的字段名、含义、单位等内容说明信息表。

(1) 系统的可控变量——设备控制参数(3 个): 冷水泵转速: chwp\_pc; 冷凝水泵转速: cwp\_pc; 冷却塔风扇转速: ct\_pc。

(2) 系统的可控变量——状态参数(12 个): 冷水泵 1, 2, 3, 4 的状态参数: chwp1stat, chwp2stat, chwp3stat, chwp4stat; 3 冷凝水泵 1, 2, 3 的状态参数: cwp1stat, cwp2stat, cwp3stat; 冷却装置 1, 2, 3 的状态参数: ch1stat, ch2stat, ch3stat; 冷却塔 1, 2 的状态参数: ct1stat, ct2stat。

(3) 其他由传感器采集数据均为不可控参数, 这些数据受到设备的配置与可控制参数以及室外温度和湿度等因素的影响, 并且直接影响系统的耗电量、冷却负载和系统效率。 注意: (1) 问题的数据均为系统实测数据, 由于某些时候受外部环境、其他电器设备、人员流动等因素的影响, 可能会出现传感器采集数据有些不稳定情况, 或个别的异常数据。(2) 在系统的调控过程中, 虽然个别设备处于关闭状态, 但会有处于待机空转而出现耗电的情况。

---

### 三、要研究的问题

附件 1 中的数据来源于热带地区某城市的一套中央空调系统,该城市常年平均温度为 25 至 32 摄氏度之间,平均湿度为 85%左右。这些数据都是根据外部环境条件和经验由人工设定的控制策略采集得到的。请你们团队充分利用附件 1 所给的数据,结合实际分析研究下列问题: 1. 对附件 1 中实际数据做深入的分析处理,挖掘出数据变化的特征和规律,研究冷却负载、系统效率、耗电量与可控变量和不可控变量之间的关系模型,并检验模型的可靠性。 2. 按照附件 1 的数据中给出的时间、室外温度和湿度、冷却负载和设备状态信息,试给出通过调节可控变量(冷水泵转速、冷凝水泵转速和冷却塔风扇转速)的最优控制策略,并给出相应的系统总耗电量和系统效率。 3. 按照附件 1 的数据中给出的时间、室外温度和湿度、冷却负载和可控变量(冷水泵转速、冷凝水泵转速和冷却塔风扇转速)取值,试给出所有设备状态变量的最优控制策略,以及相应的系统总耗电量和系统效率。 4. 按照附件 1 的数据中给出的时间、冷却负载、室外温度和湿度,试给出所有可控变量的最优控制策略,以及相应的系统总耗电量和系统效率。这里对系统设备状态变量有如下要求: (1) 每台设备的开/关以后 2 个小时以内不可以关/开; (2) 每台设备每天最多只能开关 6 次; (3) 每台设备不能超负荷运转; (4) 任何时间都至少要开启一台设备。

附件 1: 中央空调系统的采集数据表(CACS\_data.csv)

附件 2: 关于数据的字段名、含义、单位等内容说明信息表

## 二、问题分析

本题型属于多变量非线性回归和全局优化问题,首先对数据进行降维,找出有关键影响的自变量,对于相互有影响的自变量可以用表达式表示出来。

观察后可总结出:

**总耗电量=各部分的功率之和=冷却装置功率和+冷水泵功率和+冷凝水泵功率和+冷却塔功率和。**

**系统效率=总耗电量/系统冷却负载**

**各部分的平均效率=(各部分功率/冷却负载) 之和**

由于时间的影响体现在温度和湿度等方面,所以先排除时间变量。由于相对湿度对于干球温度和湿球温度有一定的函数关系,所以可以用相对湿度表示干球温度和湿球温度,而主要考虑相对湿度对因变量的影响。

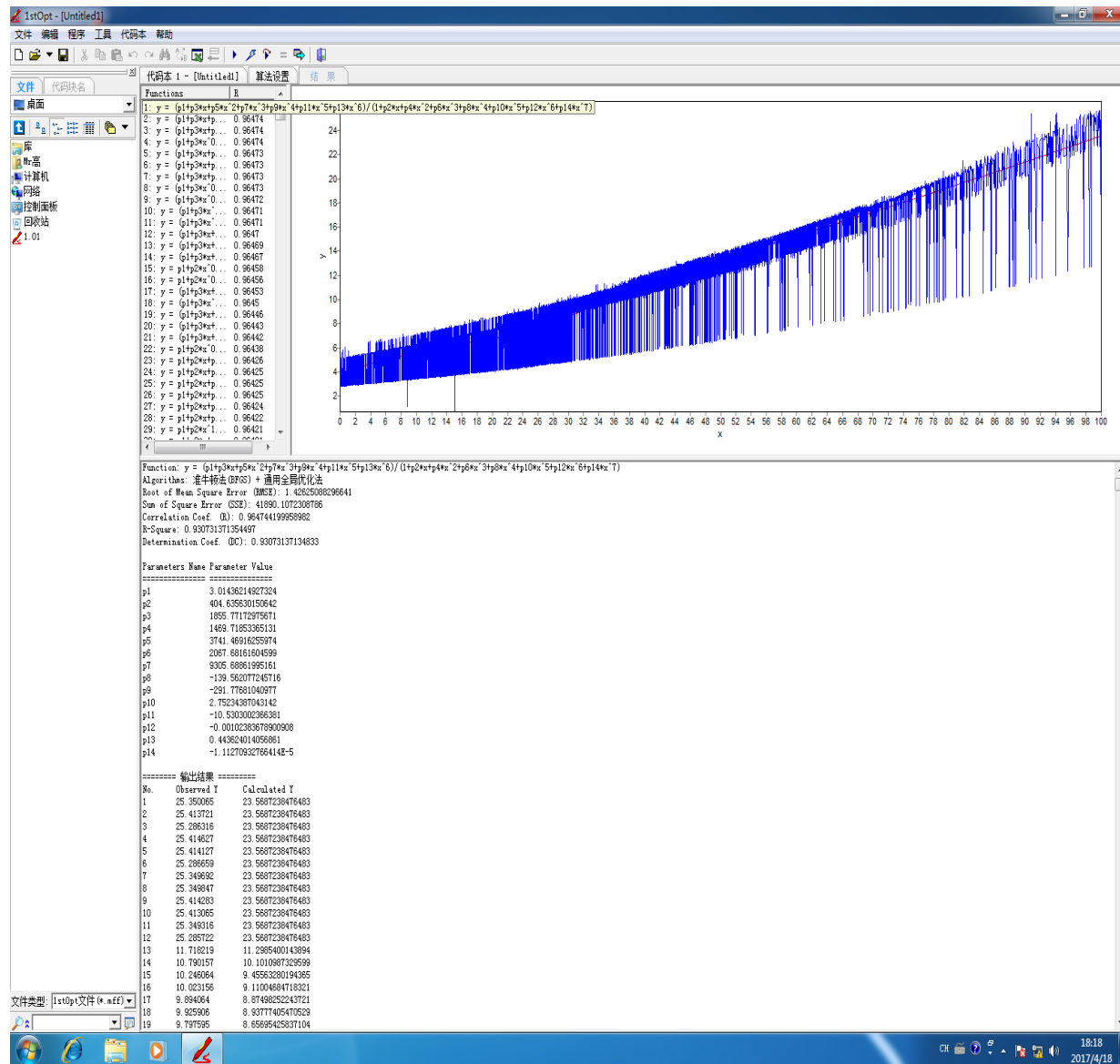
此外,冷却装置的流速,冷水泵转速,冷凝水泵转速,冷却塔风扇转速,对各自的功率起决定性影响。而冷却装置流速对冷却装置水温差有一定的函数关系,冷凝水泵的转速对其流速和温差有一定的函数关系,故可先不考虑。

即先用函数表示各部分功率,即可解出所有变量间的关系。

### 三，模型的建立和解决

运用 1stopt 软件，拟牛顿算法，可得出各部分的函数关系

1，冷凝水汞的功率和转速的 **p-v** 曲线，纵轴为功率，横轴为转速：



由图可得

$$y = (P1 + P3X + P5X^2 + P7X^3 + P9X^4 + P11X^5 + P13X^6) / (1 + P2X + P4X^2 + P6X^3 + P8X^4 + P10X^5 + P12X^6 + P14X^7)$$

均方根误差 (RMSE) : 1.42625088296641

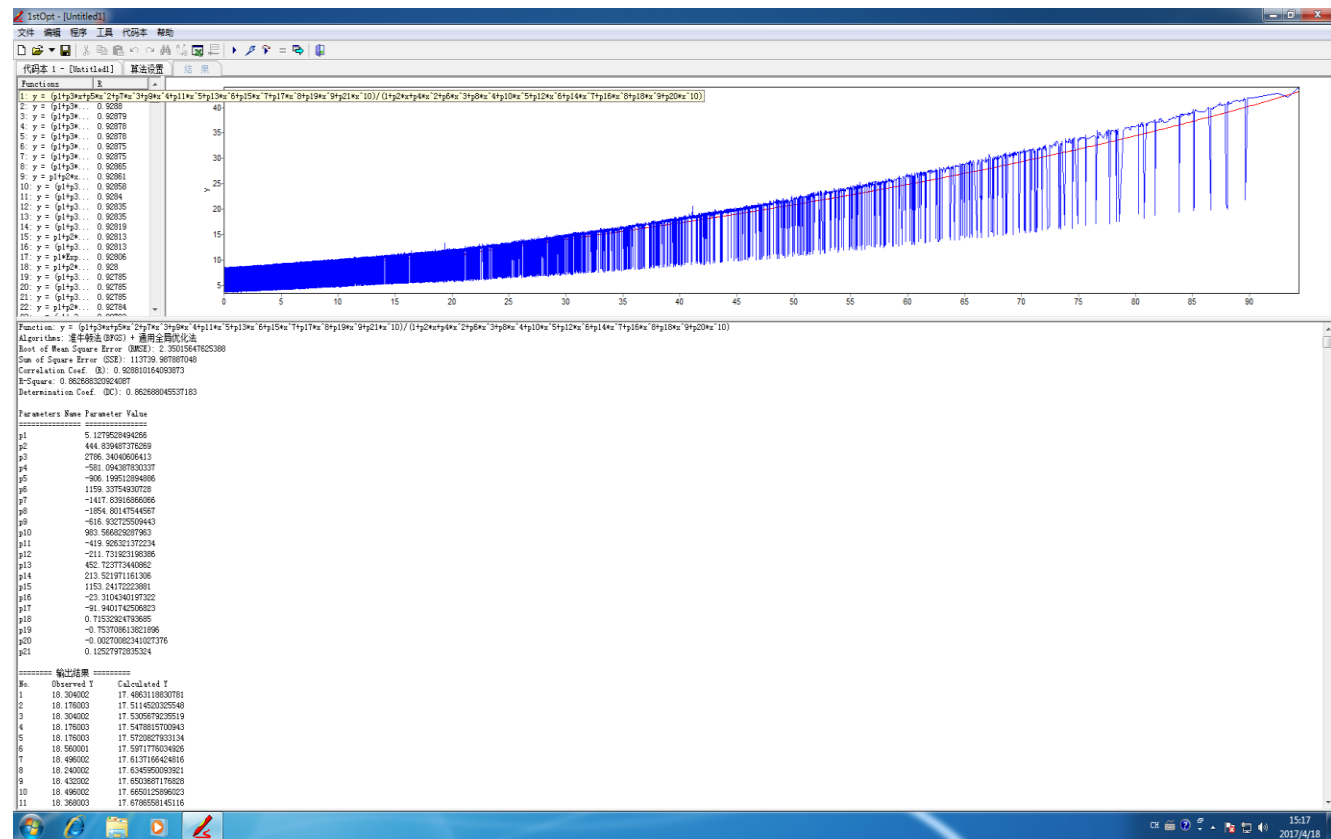
误差项平方和 (SSD) : 41890.1072308786

相关系数 R : 0.964744199958982

复相关系数 : 0.930731371354497

决定系数 : 0.93073137134833,

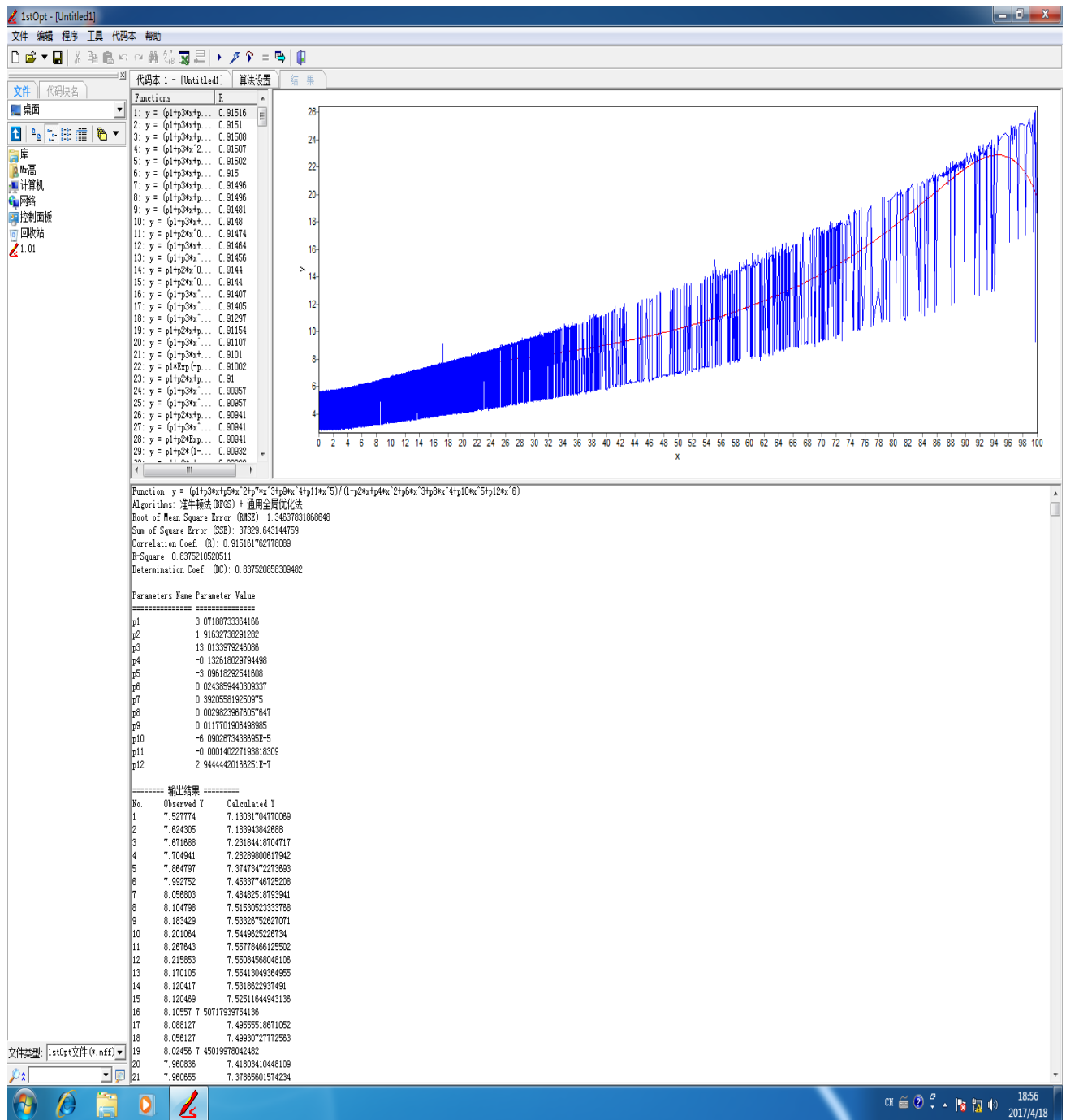
## 2.冷却塔的功率和转速 p-v 曲线:



方程和误差如图

均方根误差 (RMSE): 误差项平方和 (SSD)      相关系数 R      : 复相关系数  
 数      : 决定系数      :

### 3. 冷水泵的功率和转速 p-v 曲线:

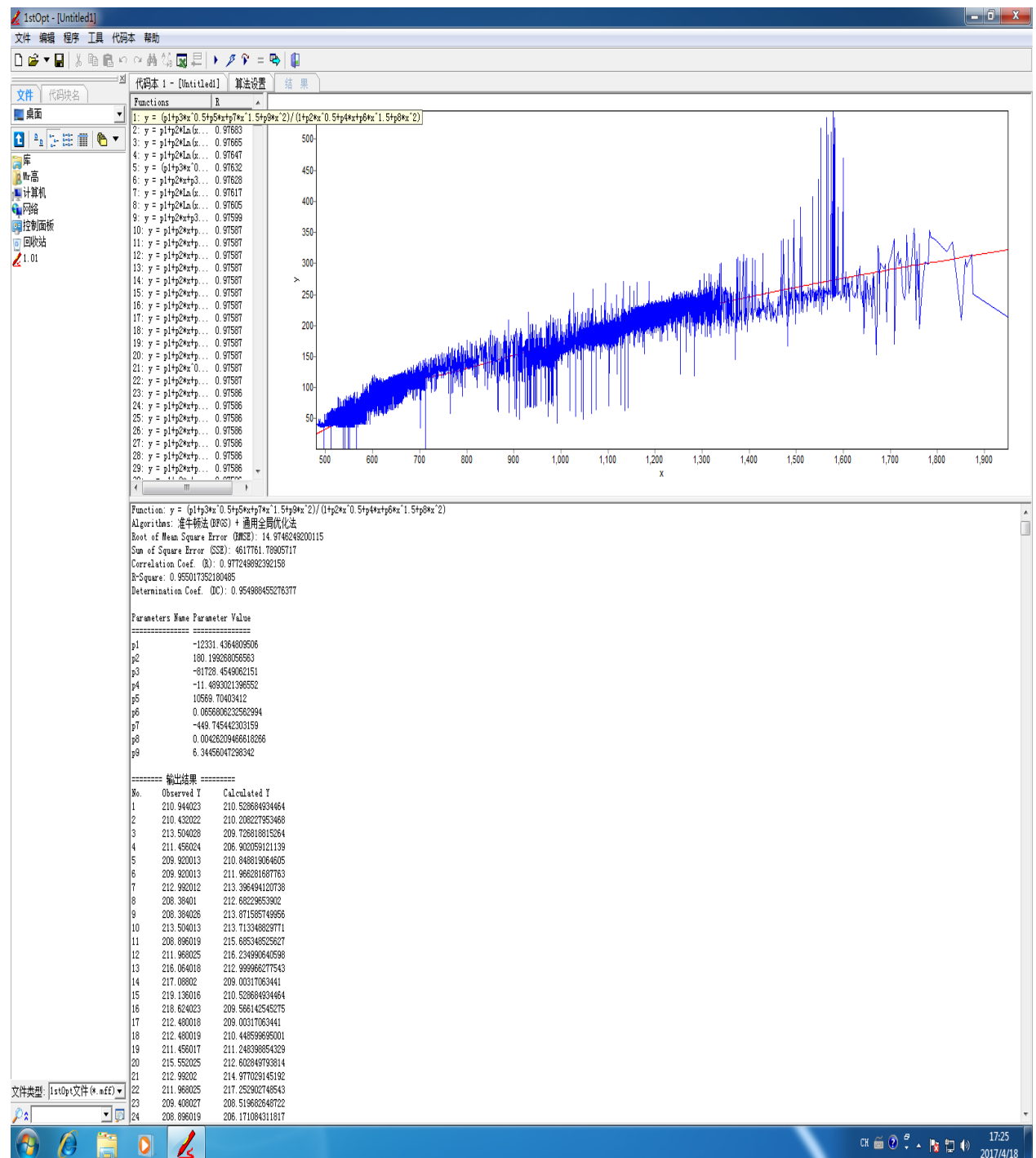


方程和误差如图

均方根误差 (RMSE): 误差项平方和 (SSD) 相关系数 R : 复相关系数  
数 : 决定系数 :



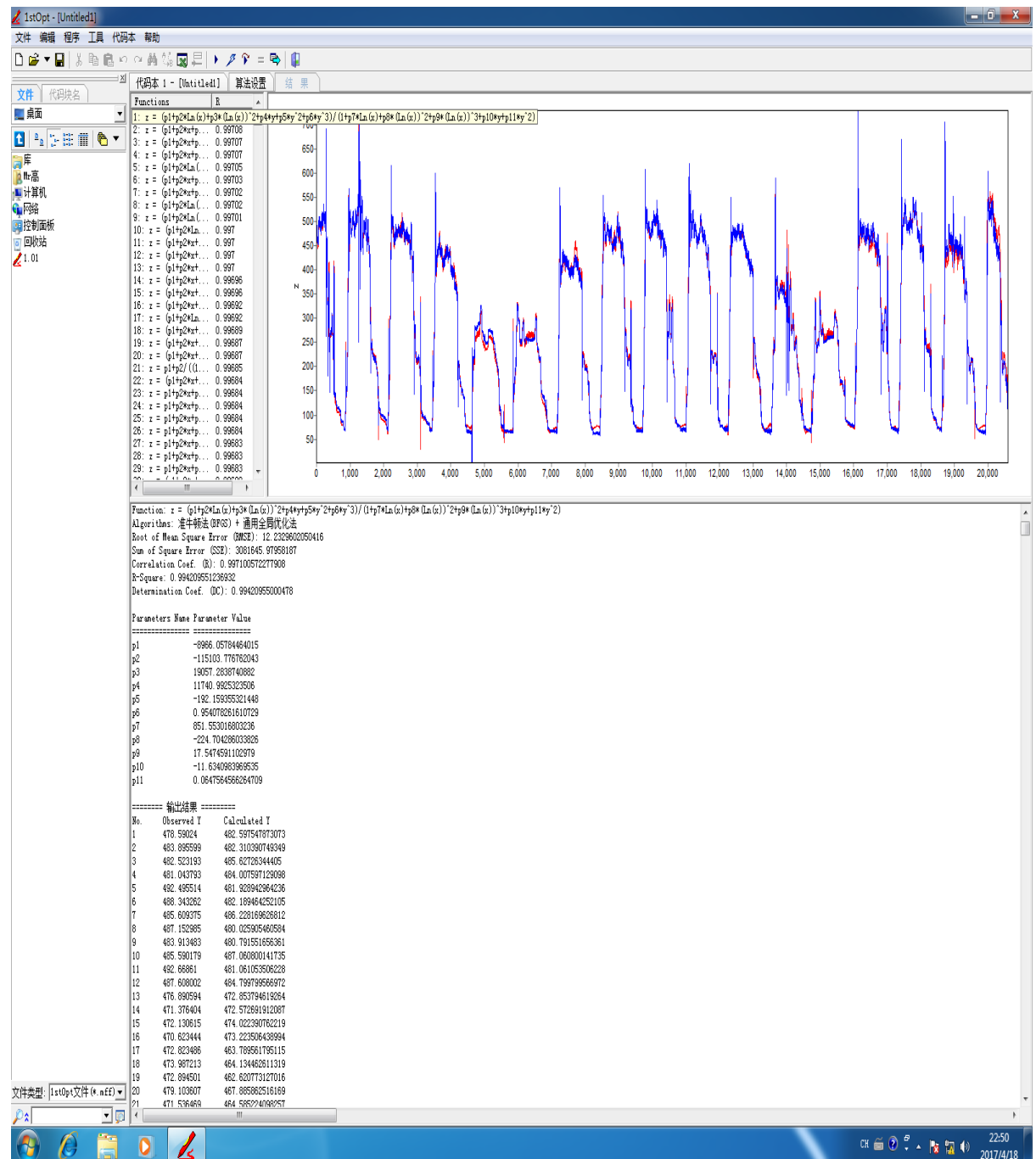
#### 4.冷却装置的功率和流速 p-v 曲线;



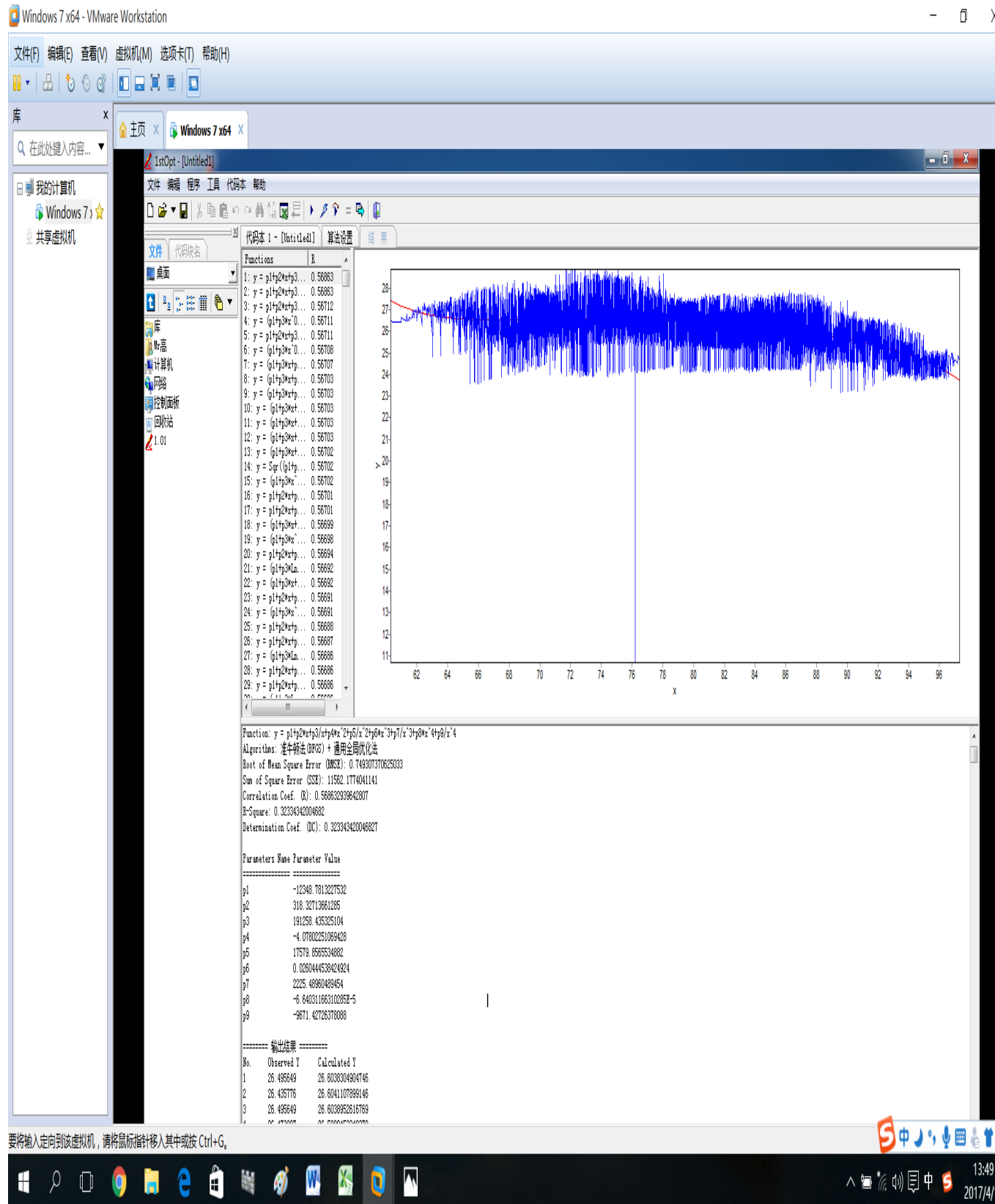
方程和误差如图

均方根误差 (RMSE): 误差项平方和 (SSD) 相关系数 R : 复相关系数  
数 : 决定系数 :

## 5.总耗电-相对湿度—负载 曲线



## 6 湿球温度—相对湿度



## 7 干球温度-相对湿度

