

# 基因表达式算法的民航涡扇发动机可调放气阀门工作基线建模

谭 燕

(中国民航飞行学院 航空发动机维修培训中心, 四川 广汉 618307)

**摘要:** 如何保障民航在运输过程中飞机发动机能够安全运行, 是航空公司都面临的难题, 因此掌握航空发动机工作阀门的特性指标, 科学地对发动机阀门进行参数的建模分析, 从而对发动机进行实时地监控就越来越重要; 针对航空发动机工作基线难以从原厂家获取的问题, 采用基因表达式编程算法 (GEP), 利用飞机快速存储记录器 (QAR) 数据对 CFM56-7B 发动机 (当前民航最主流的发动机) 的可调放气阀门 (VBV) 的工作基线进行函数挖掘分析; 结果表明, 基因表达式编程在处理阀门的非线性回归分析时, 非常接近其运行规律, 具有偏差小、精度高的优势, 并且能得到具体、方便使用的函数关系, 可以有效地提高发动机阀门监控和故障诊断的准确性。

**关键词:** 可调放气阀门; 基因表达式编程算法; 快速存储记录器; 多参数基线回归

## Operation Baseline Modeling of Civil Aviation Turbofan Engine's Variable Bleed Valve by Gene Expression Programming Algorithm

Tan Yan

(Aero Engine Maintenance Training Center, Civil Aviation Flight University of China, Guanghan 618307, China)

**Abstract:** How to ensure the safe operation of civil aviation engine during flight is the challenge for all civil airlines. Therefore, it becomes more and more important to master the characteristic index of engine's operation valves, to model and analyze the parameters of engine's operation valves scientifically and as a result, to monitor engine in real time. Aiming the problem that it is difficult to obtain the operation baseline from engine manufacturer, adopting gene expression programming algorithm (GEP), using the quick access recorder (QAR) data of CFM56-7B engine that is the most mainstream civil engine, this paper mines and analyzes the engine's variable bleed valves (VBV) operation baseline. The results show that, during VBV nonlinear regression analysis, GEP has the advantages of little deviation and high precision. GEP can also get specific and convenient function relations, which can improve the accuracy of VBV monitoring and fault diagnosis effectively.

**Keywords:** variable bleed valve; gene expression programming algorithm; quick access recorder; multi parameter baseline regression

## 0 引言

对于航空发动机来讲, 压气机的喘振会导致强烈的机械振动和热端超温, 在极短的时间内就可能造成发动机部件的严重损坏。现役的民航涡扇发动机多采用可调放气阀门 (variable bleed valve, VBV) 放出部分压气机空气以达到防喘的目的, 所以可调放气阀门的工作状态对飞行安全的影响很大。相关研究结果表明, 在飞机起飞和巡航段, 如果 VBV 阀门的开度过大, 有可能导致飞机飞行推力下降超过 10%, 引起飞机发动机推力失控事件的重大事件; 在巡航和慢车速状态时, 如果 VBV 阀门开度过小, 易引起发动机的增压机喘振故障; 在起飞、巡航和慢车状态中, 如果 VBV 阀门开度振荡不定, 会导致发动机推力下降超过

10%, 并且易发生喘振故障现象。因此研究可调放气活门对于发动机监控, 从而及时发现不良变化预防飞行事故具有重大的意义。

另一方面, 随着我国航空事业的发展, 航空发动机已不再依赖统一定期以及事故后的维修方式, 而是转变到根据每一架航空飞机的实时状态来定点视情维修。而这种更为精确的维修方式, 要求对飞机每一个部件工作参数是否正常地了解, 可以以较少的经济时间成本换取更可靠的飞行过程。然而国内所用的民用涡扇发动机基本全部由国外公司生产, 其部件特性、结构参数与控制特性等均无从得知, 因此无法建立对发动机进行视情监控和故障诊断的数学模型<sup>[1]</sup>。

近年来, 我国航空公司开始利用飞机快速存储记录器数据 (QAR) 数据实施飞机性能监控。QAR 数据包括用于飞行状态性能判断的各个设备实时温度、阀门开度、飞行速度、旋转速度等 100 多个参数。同时, QAR 译后数据库还有其他功能, 包括飞行航迹和飞行剖面的图形化查看以

收稿日期: 2019-10-28; 修回日期: 2019-11-27。

基金项目: 四川省教育厅科研项目 (18ZB0684)。

作者简介: 谭燕 (1975-), 女, 四川广汉人, 硕士研究生, 副教授, 主要从事民航航空发动机维修技术方向的研究。

及飞行数据导出为 Excel 文件等。因此可以通过对 QAR 数据的挖掘,即利用系统的输入输出参数对发动机的系统进行辨识,构造一个相对真实的反映对象的数学模型<sup>[1]</sup>。

本文即是根据 QAR 数据中的参数来辨识航空发动机阀门 VBV 的数学模型,也称工作基线,即处于良好工作状态的发动机在标准状态下 VBV 与其控制参数之间的函数关系。VBV 的控制参数较多,定性工作基线模型为:

$$VBV = f(N1K12, N2K25, VSV, TAT, T2.5, TLA, N1ACT, P0)$$

其中:

N1K12 为低压转子修正转速与额定转速的百分比(%);

N2K25 为高压转子修正转速与额定转速的百分比(%);

VSV 为可变静子叶片的角度(°);

TAT 外界大气温度(°C);

T2.5 为高压压气机进口温度(°C);

TLA 为油门杆角度(°);

N1 为低压转子实际转速与额定转速的百分比(%);

P0 为外界大气静压(PSIA)。

根据 CFM56-7B 发动机生产厂商 CFM 公司的文献[2],综合考虑影响 VBV 开度的各种因素,排除的次要参数,最终将 CFM56-7B 发动机的 VBV 的定性工作基线模型简化为:

$$VBV = f(N1K12, N2K25, VSV, TAT, T2.5)$$

本文辨识过程中所有参数样本均来源于实际的 QAR 数据。选择了一架 BOEING737-800 飞机左右两台发动机(这两台发动机 CFM56-7B 的装机时间均为 2015 年 5 月,CSN 均为 4200),收集了他们在 2018 年 1 月共 25 个航班,共 50 组 QAR 解码以后的数据。每次航班的数据由于飞行时间的不同,共有 7500 到 12000 条数据不等。在使用时依据了以下的原则预先对数据进行筛选<sup>[3]</sup>。

- 1) 选用发动机循环数较少的数据;
- 2) 选用发动机性能良好的数据;
- 3) 剔除有显著差异的数据。

除此之外,为了排除数据单位对最终结果的影响,对所有数据进行了归一化处理。为了简化计算,数据量进行了 4 倍压缩。最终,所用的数据经过了筛选、剔除和压缩等预处理操作,保证了数据的可靠和统一。

## 1 航空发动机阀门 VBV 工作基线系统辨识的常规方法

早在 60 年前,国外就开展了飞机工作性能与故障性诊断结合起来统一研究。我国 30 年前,也开始了同样的工作。比如,我国中国民航学院的林兆福教授就提出,可以应用正交实验设计原理及最小二乘法曲线拟合原理来建立航空发动机工作基线<sup>[4]</sup>。哈尔滨工业大学的钟诗胜教授又

进一步提出,利用多元非线性回归的方式,采用高斯牛顿迭代法对多元非线性回归方程进行回归分析<sup>[5]</sup>。再之后,中国民航大学的曹慧玲教授采用支持向量机回归算法,开始用厂家监控系统数据和飞机快速存储记录器数据两种方式对基线进行挖掘分析<sup>[6]</sup>。

对于航空发动机这样的复杂非线性系统,要对其进行系统辨识是比较困难的事情。针对这样的非线性模型,可使用的先进辨识算法有以下几类,分别介绍如下。

### 1.1 Volterra 泛函级数

Volterra 泛函级数具有鲜明的物理意义,切合非线性动力学的工程系统。但这种方法需要相当高的参数才能保证辨识结果的精度,计算比较困难。

### 1.2 NARMAX

NARMAX 模型即带外生变量的非线性自回归滑动平均模型。这种方法在实际中,可将给定的模型辨识问题简化为最小二乘估计问题,因此得到较为广泛的应用。但缺点是最近模型的精度取决于正确的模型结构。

### 1.3 ANN

神经网络算法(ANN)具有逼近任意非线性函数的能力。但训练算法、网络类型等需要实行指定,比较依赖先验知识,结果也不能提供简洁显性的知识表达。

### 1.4 模糊逻辑

模糊逻辑算法理论上可以以任意精度逼近任意一个定义在致密集上的非线性函数。但实际还有考虑算法的可操作性和模型的可解释性。

### 1.5 VSM

支持向量机 VSM 和神经网络一样,是一种新兴的通用学习工具,具有强大的非线性逼近能力。可以在有限样本的情况下,得到现有信息的最优解。但 VSM 得到的模型不能提供明确的数学表达式,难以分析系统的动稳态性能。

### 1.6 遗传算法

常规的遗传(GA)算法不需要先验知识,可以并行计算,全局寻优。可以快速高效地解决复杂高维和高非线性问题,具有非常突出的优点。后来 GA 算法发展出遗传编程(GP)算法,可以自动生成程序来解决问题。使用 GP 算法进行系统辨识,不需要对模型结构进行任何假设,可以在多目标下自动寻找适合的模型结构,同时优化模型结构和参数。

在 GA 和 GP 算法基础上,进一步发展了基因表达式编程(GEP)算法。这是一种基于生物基因结构和功能发明的一种新型自适应演化算法。具有 GA 和 GP 算法的优点,编码方式较为简单,易于进行遗传修饰的操作,并且具有结构和功能上的多样性。由于 GEP 在解决复杂的函数发现等问题上具有较好的性能,故在科学计算和商业应用等领域取得了广泛的应用<sup>[7]</sup>。

由于飞机发动机可调放气阀门 VBV 工作基线是一个多维，强非线性的复杂曲线。采用前面 5 种系统辨识方法，会遇到缺乏先验知识，数据量大，容易陷入局部最优，不能生成确定的数学表达式等种种问题。相比其他算法，GEP 算法具有以下优点<sup>[8]</sup>。

- 1) 具有强大的表达复杂问题的能力和快速高效的搜索能力，辨识结果不易陷入局部最优解。
- 2) 不依赖于先验知识，可以自动搜索所有可能的非线性模型空间。
- 3) 能够产生任意复杂度的模型结构，发现具有最优性能的动态表达式。
- 4) 还可以进行多目标进化，使目标模型可以在复杂度和精确度之间实现平衡。

因此本文采用 GEP 算法来解决飞机发动机可调放气阀门 VBV 工作基线系统辨识这一问题。

## 2 VBV 工作基线系统辨识的 GEP 算法设计

基因表达式编程的算法流程如图 1 所示。首先，随机产生一定数量的染色体个体（初始种群），然后，依据待解决的问题要求，计算个体的适应度。然后按照个体适应度进行选择，选中的个体进行有修饰的复制，留下具有新特性的后代。接下来，这些新的个体也要经历相同的发展过程：基因组的表达，计算适应度，选择和有修饰的复制，该过程重复若干代，直到发现一个优良解<sup>[9]</sup>。

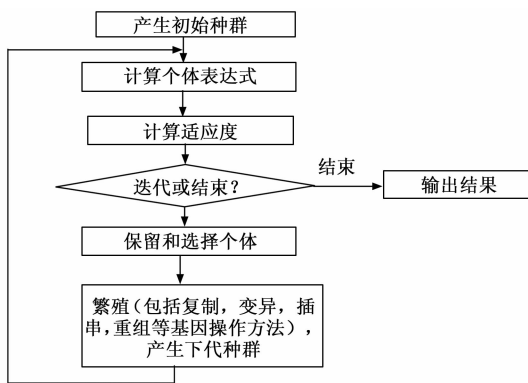


图 1 GEP 算法流程

将 GEP 算法用于工作基线函数挖掘具体工作，就是将 5 个 VBV 控制参数和可能的函数关系构造到初始染色体中，编码形成表达式树，计算表达式对应的适应度，选择适应度最好的染色体进行变异、插串、根插串、单点重组、两点重组、基因重组等操作，使染色体不断改变，越来越适应 VBV 数据所呈现的规律，直到最终匹配。适应度最高的染色体对应的表达式就是工作基线函数。

在对工作基线函数 GEP 算法编程中，所使用的开发软件是 Matlab 语言。这是美国 Mathworks 公司开发的高级技术计算语言和交互环境，广泛用于算法开发、数据可视化、数据分析和数值计算中。算法中具体设置如表 1 所示<sup>[10]</sup>。

表 1 VBV 工作基线函数挖掘中 GEP 算法相关设置

迭代次数	300
种群大小	1 000
函数集	+ , - , * , / , d/dt , sqrt() , ln() , abs() , cos() , exp()
尾部集	a , b , c , d , e
头部长度	8
染色体上基因个数	9
基因连接方法	+
染色体长度	153
变异概率	0.04
单点重组概率	0.3
两点重组概率	0.3
基因重组概率	0.1
插串概率	0.1
插串长度	1, 2, 3
根插串概率	0.1
根插串长度	1, 2, 3
适应度计算方法	SSE
精度要求	>0.95

算法中适应度的判别采用残差平方和 SSE，即拟合数据和原始数据对应点的误差平方和。计算公式如下：

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

当 SSE 的值越接近 0，说明拟合数据越接近于原始数据，模型的预测越成功，拟合度越高。以 SSE 为适应度进行判别，可以让进化过程朝着拟合优度越来越大的方向进化。以一组数据进行迭代计算的适应度变化曲线如图 2 所示。

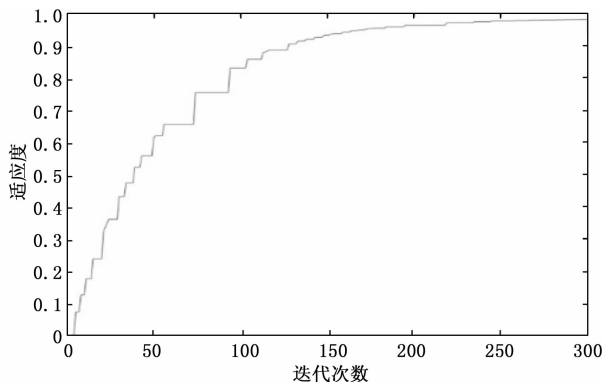


图 2 适应度变化曲线

当程序迭代到 300 代，适应度达到 0.98，此时拟合出运行曲线如图 3 (a) 所示。将拟合出的曲线图 3 (a) 和 QAR 中实际测出的曲线图 3 (b) 对比，可以看出曲线的形状基本一致，仅在少部分地方有些出入。图 3 (c) 进一步给出了拟合曲线和实测曲线之间的误差曲线，可以看出误差量很小，基本接近于零。说明 GEP 算法辨识 VBV 工作基线，误差小精度高，高效可靠。

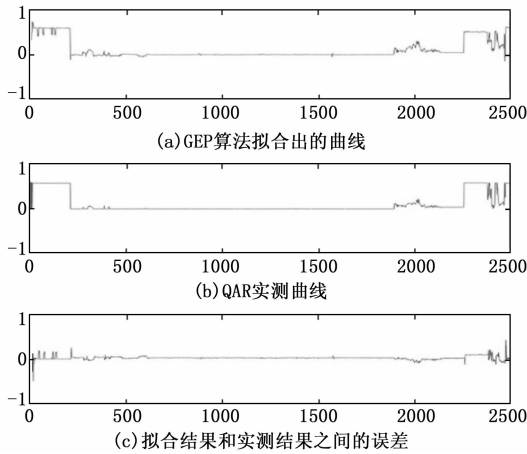


图 3 QAR 实测曲线，GEP 算法拟合曲线及两者误差

### 3 VBV 工作基线函数曲线拟合结果分析

图 3 的结果是 GEP 算法针对一个航班一台发动机的压缩数据，经过 300 代的迭代计算后，挖掘出的可调放气活门的开度 VBV 的工作规律。该规律经过表达式译码，写成公式为：

$$VBV = F(a, b, c, d, e) = 0.2957 + 4\Delta(d) + 1.15 |d^+| + 0.2ad^2 - 0.6714d - 0.0126a + 0.0043b - 0.0049c + 0.0047e$$

其中： $a, b, c, d, e$  分别代表低压转子修正转速与额定转速的百分比 N1K，高压转子修正转速与额定转速的百分比 N2K，外界大气温度 TAT，可变静子叶片的角度 VSV，高压压气机进口温度 T2.5 这 5 个控制参数。

从上述拟合公式可以看出，可调放气活门的开度 VBV 与可变静子叶片的角度 VSV 关系最为密切，公式中 2、3、4 项都和 VSV 直接有关，对 VBV 最终数值其决定性的作用。说明 VBV 主要受 VSV 数据的影响。

另外，从公式可以看出，可调放气活门的开度 VBV 除了受可变静子叶片的角度 VSV 的影响外，同时也受低压转子修正转速与额定转速的百分比 N1K，高压转子修正转速与额定转速的百分比 N2K，外界大气温度 TAT，高压压气机进口温度 T2.5 四个参数的影响，但影响相对较小。在这 4 个因素中，低压转子修正转速与额定转速的百分比 N1K 相对其它因素影响更大，在公式中第 4、6 项均和 N1K 有关。而高压转子修正转速与额定转速的百分比 N2K，外界大气温度 TAT，高压压气机进口温度 T2.5 这几项和最终可调放气活门的开度 VBV 有一定的关系，但整体影响较小。这些规律和实际中的直观影响相符，说明上述由 GEP 算法拟合出来的基线方程是贴合工程实际的。

但是该基线方程是由一次航班的一个发动机数据拟合出来的。为了深入讨论拟合结果的普遍性，将基线方程应用到其它航班数据中进行验证，验证时采用了以下指标：

确定系数 ( $R^2$ ):

$$R^2 = 1 - \frac{SSR}{SST}$$

其中：

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2, SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

确定系数  $R^2$  越接近于 1 时，说明该拟合模型的拟合优越度越高。由于确定系数会随着模型自变量数目增加而增加，为了让最终效果更精确，对确定系数进行修正，进一步采用修正确定系数 adjusted  $R^2$ 。

修正决定系数 (adjusted  $R^2$ ):

$$\text{adjusted } R^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

残差平方和 (SSE):

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

残差平方和 (SSE) 越小，拟合结果与实测数据之间的误差越小，说明拟合结果越准确。

均方误差 (MSE):

$$MSE = \frac{SSE}{N} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

均方根误差 (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

均方误差 (MSE)、均方根误差 (RMSE) 和残差平方和 (SSE) 一样，用来反映所得的拟合模型数据和实际数据之间的差异程度，数值越小说明两者越接近。

使用前面 GEP 算法拟合出来的模型公式，对随机选取的两次航班不同日期共 50 组的发动机数据进行检验。分别计算了确定系数 ( $R^2$ )、修正决定系数 (adjusted  $R^2$ )、残差平方和 (SSE)、均方误差 (MSE)、均方根误差 (RMSE)，其结果如表 2 所示。

由表 2 中结果可知，50 组数据利用函数模型进行验证所得到的结果相差不大，说明拟合函数总体可靠。确定系数 ( $R^2$ ) 和修正决定系数 (adjusted  $R^2$ ) 除去唯一一个数据以外，其余均高于 0.95 以上，说明函数的拟合优越度好，拟合数据对实测数据的解释程度高；均方误差 (MSE) 在 0.010 2 ~ 0.023 0 之间，平均 0.014 8；均方根误差 (RMSE) 在 0.105 0 ~ 0.144 3 之间，平均 0.120 8，表明拟合函数数据分散在实际曲线附近，离散度很小，与实际曲线契合程度非常高。

### 4 结束语

通过对民航涡扇发动机 QAR 数据的挖掘，利用 GEP 算法找到了一个相对真实的民航涡扇发动机可调放气活门 (VBV) 的工作基线数学函数。验证结果表明，GEP 算法挖掘得到的结果精度高，误差小，在有限样本条件下对于民航涡扇发动机 VBV 工作基线的非线性建模具有很高的准确性，所构建的模型可以满足民航对可调放气活门进行监控和故障诊断的需要。

表 1 GEP 算法拟合模型对随机选取发动机数据组检验结果

Data	$R$	$R^2$	$adR^2$	SSE	MSE	RMSE
20180107-1-1	0.987 6	0.975 3	0.975 3	113.931 5	0.011 2	0.105 9
20180107-1-2	0.989 2	0.978 5	0.978 4	118.263 1	0.011 6	0.107 9
20180107-2-1	0.981 2	0.962 8	0.962 8	144.622 6	0.017 2	0.131 2
20180107-2-2	0.987 3	0.974 8	0.974 8	144.398 8	0.017 2	0.131 1
20180108-1-1	0.987 0	0.974 1	0.974 1	155.175 2	0.011 0	0.105 0
20180108-1-2	0.986 0	0.972 3	0.972 2	170.711 7	0.012 1	0.110 2
20180108-2-1	0.985 8	0.971 8	0.971 8	137.794 1	0.011 9	0.108 9
20180108-2-2	0.986 2	0.972 6	0.972 6	141.625 4	0.012 2	0.110 4
20180108-3-1	0.979 1	0.958 6	0.958 6	120.345 4	0.013 9	0.117 8
20180108-3-2	0.983 0	0.966 2	0.966 2	117.483 1	0.013 6	0.116 4
20180109-1-1	0.976 6	0.953 7	0.953 7	114.228 6	0.016 5	0.128 5
20180109-1-2	0.978 8	0.958 0	0.958 0	114.900 3	0.016 6	0.128 9
20180109-2-1	0.989 9	0.980 0	0.980 0	129.327 7	0.011 7	0.108 2
20180109-2-2	0.991 1	0.982 2	0.982 2	130.735 0	0.011 8	0.108 8
20180109-3-1	0.987 9	0.975 9	0.975 9	135.992 2	0.016 6	0.128 8
20180109-3-2	0.990 1	0.980 4	0.980 4	129.992 3	0.015 9	0.126 0
20180109-4-1	0.987 3	0.974 7	0.974 7	135.014 0	0.012 1	0.109 9
20180109-4-2	0.987 0	0.974 2	0.974 2	141.440 7	0.012 6	0.112 5
20180109-5-1	0.984 8	0.969 9	0.969 8	134.339 2	0.016 2	0.127 4
20180109-5-2	0.967 0	0.935 1	0.935 0	172.463 6	0.020 8	0.144 3
20180110-1-1	0.982 4	0.965 1	0.965 0	113.954 7	0.013 8	0.117 3
20180110-1-2	0.982 5	0.965 2	0.965 2	114.235 6	0.013 8	0.117 5
20180110-2-1	0.970 3	0.941 5	0.941 5	121.010 4	0.016 6	0.128 7
20180110-2-2	0.974 8	0.950 2	0.950 1	114.416 7	0.015 7	0.125 2
20180111-1-1	0.987 3	0.974 9	0.974 8	119.889 8	0.013 2	0.115 0
20180111-1-2	0.986 5	0.973 3	0.973 3	117.985 0	0.013 0	0.114 1
20180111-2-1	0.961 4	0.924 2	0.924 2	159.958 3	0.023 1	0.152 1
20180111-2-2	0.962 7	0.926 8	0.926 8	159.101 6	0.023 0	0.151 7
20180112-1-1	0.984 3	0.968 8	0.968 8	134.990 0	0.011 4	0.106 9
20180112-1-2	0.984 2	0.968 7	0.968 7	146.872 4	0.012 4	0.111 5
20180113-1-1	0.983 3	0.967 0	0.966 9	126.479 9	0.011 6	0.107 7
20180113-1-2	0.988 4	0.976 9	0.976 9	122.466 4	0.011 2	0.106 0
20180113-2-1	0.986 7	0.973 5	0.973 5	121.721 8	0.011 2	0.105 7
20180113-2-2	0.989 5	0.979 1	0.979 1	126.699 6	0.011 6	0.107 8
20180113-3-1	0.981 8	0.964 0	0.963 9	80.144 1	0.014 4	0.120 0
20180113-3-2	0.982 6	0.965 6	0.965 5	84.872 5	0.015 3	0.123 5
20180113-4-1	0.952 6	0.907 5	0.907 5	161.787 1	0.016 2	0.127 3
20180113-4-2	0.978 0	0.956 4	0.956 4	144.034 8	0.014 4	0.120 2
20180114-1-1	0.982 3	0.965 0	0.965 0	118.121 1	0.010 2	0.100 9
20180114-1-2	0.978 6	0.957 6	0.957 6	123.304 6	0.010 6	0.103 1
20180114-2-1	0.984 2	0.968 6	0.968 6	127.651 6	0.011 0	0.105 0
20180114-2-2	0.986 8	0.973 7	0.973 7	137.639 1	0.011 9	0.109 0
20180114-3-1	0.984 6	0.969 5	0.969 5	123.854 3	0.010 8	0.103 8
20180114-3-2	0.983 7	0.967 8	0.967 7	123.406 1	0.010 7	0.103 6
20180114-4-1	0.983 3	0.967 0	0.966 9	126.479 9	0.011 6	0.107 7
20180114-4-2	0.988 4	0.976 9	0.976 9	122.466 4	0.011 2	0.106 0
20180115-1-2	0.967 3	0.935 7	0.935 6	146.216 7	0.011 4	0.107 0
20180115-2-1	0.984 7	0.969 6	0.969 5	130.233 4	0.012 0	0.109 4
20180115-2-2	0.987 8	0.975 7	0.975 7	130.519 9	0.012 0	0.109 5
20180115-3-1	0.987 3	0.974 8	0.974 8	138.103 6	0.010 9	0.104 5