

文章编号: 2095-2163(2022)05-0154-06

中图分类号: TK09

文献标志码: A

基于降维算法从少量测量数据中重构温度场

罗芸¹, 钱进¹, 王一桂², 朱道兴²

(1 贵州大学 电气工程学院, 贵阳 550025; 2 中国电建集团贵州工程有限公司, 贵阳 550025)

摘要: 由于测量环境及测量技术的制约,工业中温度信息的获取主要依据少量测点数据,难以得到被测对象完整温度信息。计算方法可以获得被测对象完整温度信息,但计算量大,信息再现实时性差。为了解决这个问题,实现快速、有效、准确的实时温度分布测量,提出一种基于 Gappy POD 低维表示的温度分布降维重建算法,从少量局部温度测量数据准确中重建温度场,并分析采用上述算法进行温度场重构时 POD 基个数、测点数量和测点位置对其的影响。通过双峰温度场模型重建结果和误差分析表明,重建误差在 1×10^{-10} 以下,且该算法显著地减少了测量传感器的数量,降低了温度分布测量的复杂性与测量成本,为工业测量中数据缺失提供新的解决思路。

关键词: Gappy POD; 温度场重构; POD 基; 少量测量数据; 降维

Reconstructing the temperature field from a small amount of measured data based on Gappy POD algorithm

LUO Yun¹, QIAN Jin¹, WANG Yigui², ZHU Daoxing²

(1 College of Electrical Engineering, Guizhou University, Guiyang 550025, China;

2 China Power Construction Group Guizhou Engineering Co., Ltd., Guiyang 550025, China)

[Abstract] Due to the restriction of measuring environment and measuring technology, the acquisition of temperature information in industry is mainly based on a small amount of measuring point data, so it is difficult to get the complete temperature information of the measured object. The calculation method can obtain the complete temperature information of the measured object, but the calculation amount is large and the real-time information reproduction is poor. In order to solve this problem, and realize the rapid, effective and accurate measurement of temperature distribution, a dimensionality reduction reconstruction algorithm of temperature distribution based on Gappy POD low-dimensional representation is proposed, which can reconstruct the temperature field accurately from a small amount of local temperature measurement data, and analyzes the influence on it of the number of POD bases, the number of measurement points and the measurement points position when the above algorithm is used to reconstruct the temperature field. The reconstruction results and error analysis of the bimodal temperature field model show that the reconstruction error is below 1×10^{-10} , and the algorithm significantly reduces the number of measurement sensors, reduces the complexity and measurement cost of temperature distribution measurement, and provides a new solution for data missing in industrial measurement.

[Key words] Gappy POD; temperature field reconstruction; POD base; a small amount of measured data; dimension reduction

0 引言

温度分布信息对锅炉、大气层、海洋等的监测都非常重要,其获取方法大概可以分为2种:计算方法和测量方法。在实际应用中,2种方法都有其局限性^[1-3]。计算方法是数学问题离散化形式表示,但由于计算方法运算过程复杂且对计算资源要求较高,温度信息再现实时性差。测量方法受测量环境、被测物结构、测量设备安装以及工程造价等因素的影响,难以获得全面温度分布信息。

为了对适合工业现场温度分布进行快速、准确

的重建,将降维算法引入温度分布重建,综合计算方法和测量方法的优点,由计算方法提供温度分布重建的样本数据集,测量方法得到少量温度传感器数据,使用数据降维算法提取样本数据集中温度分布特征,结合有限温度测量数据对温度分布实现实时重建。

利用本征正交分解(POD)降维方法进行温度分布重建以及分析,获得良好的效果,也是近些年来研究热点^[4-7]。POD算法是数据驱动的维度降低算法,从大量的已知数据中提取元数据的主要特征,得到原数据的低维表达。Woojin 等人^[8]采用本征

基金项目: 贵州省科技支撑计划项目([2020]2Y040)。

作者简介: 罗芸(1997-),女,硕士研究生,主要研究方向:热动力过程数值模拟与仿真;钱进(1965-),男,副教授,硕士生导师,主要研究方向:热动力过程多尺度数值模拟与仿真。

通讯作者: 钱进 Email: 1320064076@qq.com

收稿日期: 2021-11-22

正交分解方法提取低维基向量, 重构了 500 MW 切向燃烧煤粉锅炉温度场, 重构效果良好。在 POD 方法的基础上 Sirovich 等人^[9-10]首次提出了 Gappy POD 方法, 该方法利用 POD 基结合部分测量数据重建缺失数据进行补全重构。与常用的基于 POD 的降维方法相比, Gappy POD 的显著优点之一是其系数矢量是根据部分测量数据求解的, 不需要了解测量对象物理过程的详细信息。Lei 等人^[5]基于 Gappy POD 方法, 提出数值模拟信息和测量信息相结合, 从局部测量数据中重建稳定温度场, 通过数值仿真验证了其可行性和有效性。陈敏鑫等人、及孙单勋^[7, 11-12]提出了一种将 CFD 信息融合至 Gappy POD 算法中实现了对物理场的实时重建方法, 该方法重建范围大, 准确性高, 为未来物理场预测重建提供了一种新的思路。

综上所述, 提出基于 Gappy POD 算法, 结合部分测量数据对温度场进行准确重建, 具体重建流程如图 1 所示, 并分析研究 Gappy POD 算法进行重建时, POD 基个数、传感器数量和测量位置等因素对温度场重建的影响。

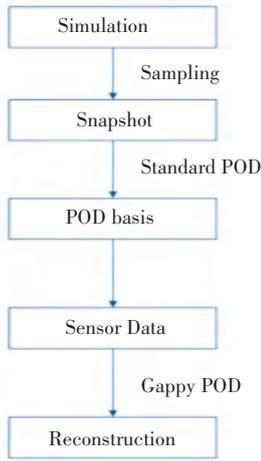


图 1 重建流程

Fig. 1 Reconstruction process

1 本征正交分解方法

1.1 标准 POD 方法

标准 POD 方法主要思想是将原始温度场分解为基函数 (POD 基) 和基系数的线性组合。应用 POD 算法, 原始温度场组成的快照矩阵的大部分信息由少数 POD 基近似表示, 进而实现了降维。

将多组温度场数据构成快照矩阵合集 $\mathbf{T}(x_i, t_j)$ ($1 < i, j < N$), x 是空间坐标, t 是温度场标量编号。其样本矩阵形式可表示为:

$$\mathbf{T} = [T_1, T_2, \dots, T_i] \quad (1)$$

计算快照矩阵各个节点的平均值, 即:

$$\mathbf{T}(x_i) = \frac{1}{N} \sum_{j=1}^N \bar{\mathbf{T}}(x_i, t_j) \quad (2)$$

由此得到温度场的脉动量矩阵, 即:

$$\hat{\mathbf{T}}(x_i, t_j) = \mathbf{T}(x_i, t_j) - \bar{\mathbf{T}}(x_i) \quad (3)$$

计算 $\hat{\mathbf{T}}(x_i, t_j)$ 的相关矩阵 \mathbf{R} , 并求其特征值和特征向量, 即:

$$\mathbf{R} = \frac{1}{N} \hat{\mathbf{T}}^T \hat{\mathbf{T}} \quad (4)$$

$$\mathbf{R}\mathbf{A} = \lambda\mathbf{A} \quad (5)$$

其中, λ 为特征值, \mathbf{A} 为特征向量。

通过式 (6)、式 (7) 可以计算各阶 POD 基 $\Phi_j(x)$ 和其对应的模态系数 $a_j(t)$, 具体公式如下:

$$\phi_j(x) = \frac{1}{\sqrt{\lambda_j}} \hat{\mathbf{T}}\mathbf{A}_j \quad (6)$$

$$a_j(t) = \frac{\Phi_j(x)^T \cdot \hat{\mathbf{T}}}{\Phi_j(x)^T \cdot \Phi_j(x)^T} \quad (7)$$

POD 基表示捕获温度场的主要特征, 前 m 个 POD 模态所捕获的能量占全阶模态的能量为:

$$E = \sum_{j=1}^m \lambda_j / \sum_{j=1}^k \lambda_j \quad (8)$$

快照矩阵中任意温度场可以由温度场的平均值和一组基模态和线性组合来重构, 即:

$$\mathbf{T}(x, t) = \bar{\mathbf{T}}(x) + \sum_{j=1}^N \phi_j(x) a_j(t) \quad (9)$$

1.2 Gappy POD 方法

根据相关文献^[5-6, 13-19], 将含有 Gappy 数据的目标场 $\hat{\mathbf{T}}(x)$ 定义为一个掩码向量 \vec{n} 和一个完整目标场 $\mathbf{T}(x)$ 的点乘:

$$\hat{\mathbf{T}}(x) = \vec{n} \cdot \mathbf{T} = \sum_{i=1}^m \vec{n}^i \mathbf{T}(x_i) \quad (10)$$

掩码向量 \vec{n} 由 0 和 1 组成:

$$\vec{n}^i = \begin{cases} 0 & \text{if } \hat{\mathbf{T}}(x) \text{ lost} \\ 1 & \text{if } \hat{\mathbf{T}}(x) \text{ not lost} \end{cases} \quad (11)$$

对于重建目标场 $\hat{\mathbf{T}}(x)$ 可以表示为:

$$\hat{\mathbf{T}}(x) = \sum_{k=1}^m b_k \phi_k(x) \quad (12)$$

其中, ϕ_k 表示 $\hat{\mathbf{T}}(x)$ 的归一化的 POD 的基向量, b_k 表示未知系数。

实际中, ϕ_k 是利用标准 POD 方法从温度矩阵

$T(x_i, t_j)$ ($1 < i, j < N$) 中提取出来的基向量。所以,存在部分测量数据时,对于温度矩阵中的每一个样本,可以计算完整目标场 $T(x)$ 和重建目标场 $\tilde{T}(x)$ 的误差为:

$$e = \|T - \tilde{T}\|_2 = \|T - \sum_{k=1}^k b_k \phi_k\|_2 \quad (13)$$

为了使 e 值最小,对上式中的 b_k 求偏导,令偏导等于 0 得到误差最小值。研究推得的数学公式如下:

$$\frac{\partial e}{\partial b_k} = 0 \quad (14)$$

整理得:

$$\vec{M} \cdot \vec{b} = \vec{F} \quad (15)$$

$$M_{ij} = (\phi_i, \phi_j); F_j = (T, \phi_j)$$

进一步地,可以计算出:

$$M = (\vec{n}\phi)^T (\vec{n}\phi) \quad (16)$$

将式(16)结果带入式(12)就可以得到重建目标场 $\tilde{T}(x)$ 。利用重建目标场 $\tilde{T}(x)$ 可以补全含有缺失 Gappy 数据得目标场 $\hat{T}(x)$ 。

为了后文的论述,定义均方根误差为重建误差:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (T(x_i) - \tilde{T}(x_i))^2} \quad (17)$$

2 温度场重建仿真研究

2.1 二维温度场模型仿真

为了对待测对象进行二维温度场仿真研究,采用双峰偏斜温度场模型,运用 Matlab 软件进行仿真实验。

双峰偏斜数学模型:

$$T(x, y) = a \times \sin\left(\frac{\pi}{4}x\right) \cdot \sin\left(\frac{\pi}{4}y\right) + b \times e^{-30 \times \left(\frac{x}{4.5} - \frac{3}{4}\right)^2 - 35 \times \left(\frac{y}{3.5} - \frac{2}{3}\right)^2} + a \times e^{-25 \times \left(\frac{x}{2} - \frac{4}{7}\right)^2 - 50 \times \left(\frac{y}{4} - \frac{2}{5}\right)^2} + 600 \quad (18)$$

其中, $T(x, y)$ 表示双峰偏置温度分布模型坐标点 (x, y) 处温度值(K); x 表示横坐标(m); y 表示纵坐标(m); a, b 表示边界条件。

使用式(18)在计算区域内计算温度值,计算区域为 $4 \times 4 \text{ m}^2$, 采样步长为 0.1 m , 采集 1 681 个温度值,形成温度分布数据。计算不同边界条件下的温度分布数据,构建样本数据集以及测试集。设定双峰偏斜温度分布模型样本数据集以及测试集边界条件见表 1、表 2。

表 1 样本数据集边界条件

Tab. 1 Boundary conditions of sample data sets

边界条件 a	边界条件 b
500	600
600	700
700	800
800	900
900	1 000
1 000	

表 2 测试数据集边界条件

Tab. 2 Boundary conditions of test data sets

工况名称	边界条件 a	边界条件 b
工况 1	523	784
工况 2	685	924
工况 3	750	850
工况 4	394	406
工况 5	2 976	3 125

根据表 1、表 2 给出的边界条件,共有 30 组样本数据和 5 组测试数据,应用公式(15),计算温度分布数据,采样形成维数为 $1\ 681 \times 30$ 的样本数据集和维数为 $1\ 681 \times 5$ 的测试数据集。

2.2 温度场重建结果

在本节中,使用 Gappy POD 方法结合部分温度测量数据重建温度场。将样本数据集进行特征分解,按照公式(8)确定 POD 基的个数,构成转化矩阵,选择几个坐标点作为实际温度测点,坐标点的温度数据作为温度测量值,结合转化矩阵对测试数据集进行温度场重建。

首先对 30 个不同工况的样本数据集进行 POD 分析,得到用于重构温度场的 POD 基。不同 POD 基所占能量曲线和前两阶 POD 模态分别如图 2、图 3 所示。从图中可以看出前两阶 POD 模态所占能量几乎是总能量的 100%,这意味着高维数据可以用 2 个空间模态来准确表达,高维数据可以得到显著压缩。

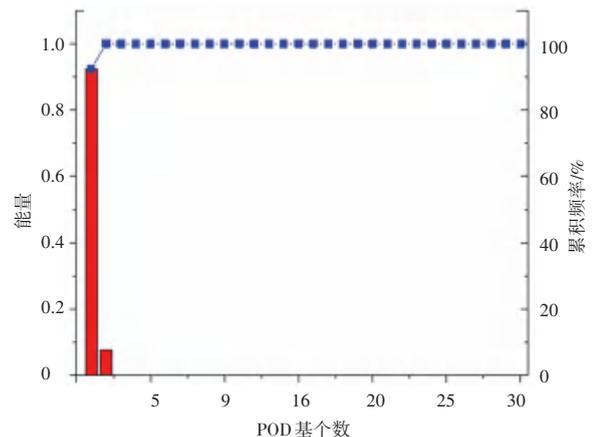


图 2 各阶 POD 基所占能量

Fig. 2 The energy of each order POD base

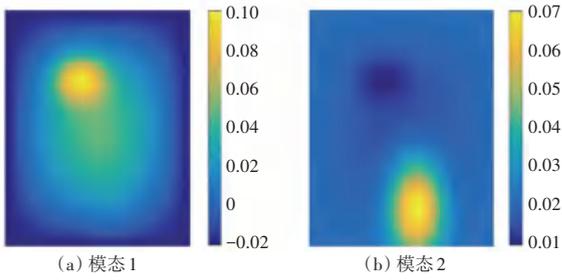


图 3 前两阶 POD 模态

Fig. 3 The first two order POD

其次,对表 2 中的 5 组测试工况选择前两阶 POD 模态、10 个温度测点和相同测点位置进行温度分布重建。重建结果如图 4 所示,图 4(a)~(e)展示了测试工况原始温度场,图 4(f)~(j)给出了重建温度场结果。按照公式(17)计算重建误差如图 5 所示。

由图 4 可以看出,重建温度分布与原始温度分布大致保持一致,重建效果良好。

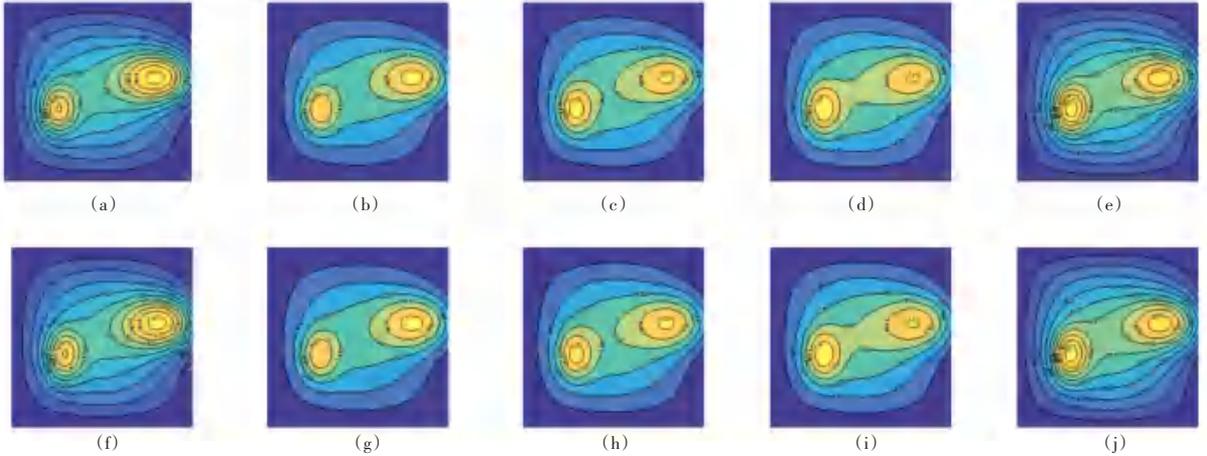


图 4 重建结果

Fig. 4 Reconstruction results

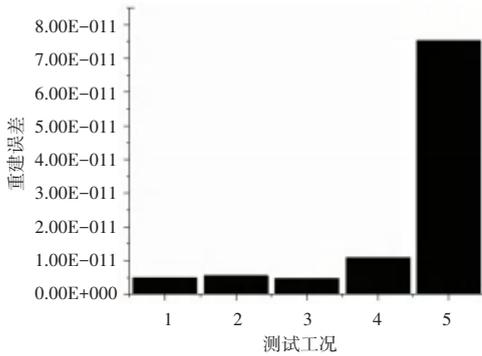


图 5 测试工况重建误差

Fig. 5 Test condition reconstruction errors

图 5 中,样本工况范围内的测试工况 1~3 的重建误差小于在样本范围外的测试工况 4~5 的重建误差,但其温度分布重建效果同样良好,全部测试工况重建误差均在 1×10^{-10} 之下;测试工况 4 离样本工况范围近,测试工况 5 离样本工况范围远,测试工况 5 的重建误差大于测试工况 4 的重建误差。因此结合 Gappy POD 和部分测量数据的温度分布重建算法有着良好的适用性和准确性。

3 温度场重建结果误差分析

针对温度场的重建,使用控制变量法来分析

POD 基个数、传感器数量(测点数量)、传感器放置位置(测点位置)等因素对重建结果造成的影响。为了消除实验带来的随机误差,所有结果均进行了多次,做图数据为多次实验的平均值。

3.1 POD 基数量对重建结果的影响

图 6 中展示了测试工况 1 的重建误差随着 POD 基数量的变化情况。利用 10、50 和 70 个实际测量数据,在固定的测量数据情况下,随机选取不同位置的测点进行温度重建计算,做图数据为多次计算的平均值,消除测点位置对结果产生的影响。从图 6 中可以看出,使用前两阶 POD 基进行重构,比使用前一阶 POD 基进行重构,重构误差大大减小,重构误差随着 POD 基数量的增加,先减小、而后趋于稳定。3 种不同的测点数据下,重建误差随 POD 基数量的变化保持相同的趋势。

3.2 测点位置对重建结果的影响

图 7 为 3 和 5 个传感器在随机位置点的重构误差。

从图 7 中可以看出,由于测量数据的测点位置是随机的,即使同样数量的测点数据,测点位置不同,重建误差变化较大。

图 8 为利用 2 个 POD 基在不同传感器数量和

不同随机位置下重构的误差,选取不同数量的测量数据,随机选择测点位置,进行5组实验,得到5组计算结果。由图8可以看出,不同随机位置测点的重建误差均是随传感器测量数据数量的增加由降低到趋于稳定,但测点位置不同,其重建误差也不相同。

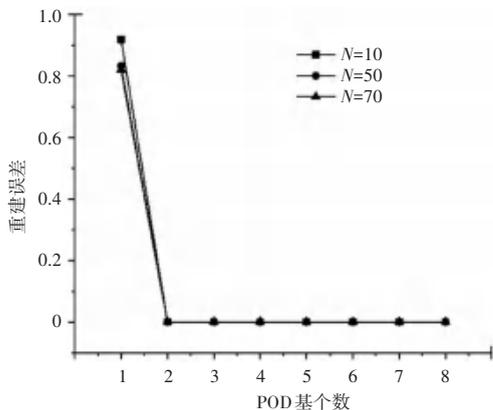


图6 利用10、50和70个测量数据,重建误差随POD基数量变化曲线
Fig. 6 The reconstruction errors curve with the number of POD bases by using 10, 50 and 70 measurements

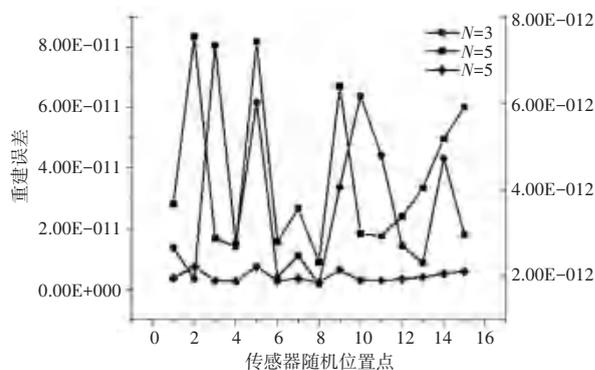


图7 3和5个传感器数量在随机位置点下的重建误差

Fig. 7 Reconstruction errors of 3 and 5 sensors at random position points

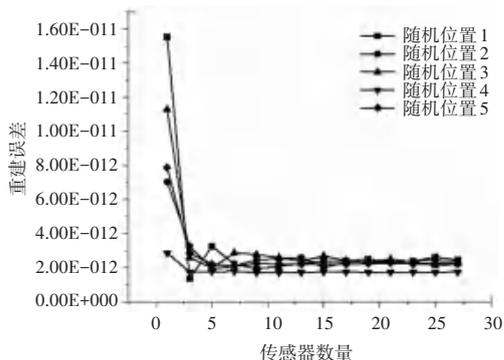


图8 不同传感器数量和不同随机位置的重建误差

Fig. 8 Reconstruction errors of different random positions with different number of sensors

由此可见传感器的放置位置十分重要,不同的放置位置对其在温度分布重建算法中影响程度也不同。

3.3 测点数量对重建结果的影响

在传统的温度预测和重构中,预测和重构结果准确度高需要大量的传感器测量数据做支撑,然而,大量传感器使用却需要大量资金、人力和物力投入。因此,尽量少的传感器数量使用也是衡量预测重构模型的标准之一。由上文可知测点位置对重建结果的影响,为消除此影响,对此进行多次计算取平均值。

利用一阶、二阶和三阶POD基进行重建,重构误差如图9所示。由图9可以看出,一阶POD基重建误差随传感器数量增加虽呈下降趋势,但过程跌宕起伏;二阶和三阶POD基重建误差随传感器数量增加而减少、再趋于稳定。一阶与二、三阶趋势不相同,这是因为一阶POD基所占能量较低,对数据反应能力较弱,这与前文中提出的结论是一致的。

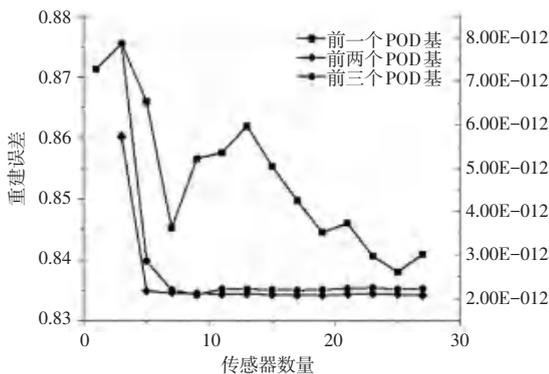


图9 不同POD基在不同传感器下的重构误差

Fig. 9 Reconstruction errors under different sensors with different POD bases

由图8和图9可知,重建误差均在测点数量为5后趋于稳定,由此可知传感器最佳数量为5个。

4 结束语

(1) 针对有限的已知测量数据的温度场重构,本文结合Gappy POD方法,利用少量POD基能够快速准确地重构温度场,为了验证算法的可行性和适用性,重建5组测试工况,重建误差均在精度要求范围内,对于其边界条件不在快照矩阵范围内的测试工况4、5的重建,仍有令人满意的重建结果。

(2) 重构误差随POD基数量的增加先减少、而后趋于稳定,少量的POD基就能重构温度场,对后续低阶模型有一定指导意义。

(3) 同样数量的传感器,放置位置不同,重建误差变化较大,由此可见测点位置所携带信息重要程度在重建算法中也各有不同,因此测点位置的优化布置尤为重要,后期会对最佳测点位置方式进行研究。

(4) 重建误差随传感器数量的增加先减少、再