DOI:10.13203/j.whugis1988.02.001

第13卷 第2期 武汉测绘科技大学学报 Vol.13 No.2
1988年 6月 Journal of Wuhan Technical University of Surveying and Mapping Jun. 1988

主分量分析与数学模型扭曲法一致性的研究*

王 建 国

摘 要

本文研究测量控制网的主分量分析与数学模型扭曲法之间的一致性关系，以高能物理加速器的环形网为对象进行了模拟试验分析，结果表明，在一定情况下，二者是一致的。这不但为揭示模拟曲线存在特定分布规律的内在原因提供了严密的理论依据，也为避免大量模拟试验找到了一条有效的途径。

【关键词】 主分量分析；数学模型扭曲法；环形网；点位径向主分量曲线；点位径向位移模拟曲线

1 問 題 的 提 出

数学模型扭曲法（〔5、6〕和〔7〕等）就是根据模拟的观测误差，研究测量控制网中偶然误差的不同分布对待定参数及其函数的影响，即偶然误差在待定参数及其函数上的分布规律。该方法已作为精密工程控制网优化设计的方法之一，引起国内外的重视。苏联谢尔普霍夫高能物理加速器[8]所布设的环形直伸导线网（参见图1），采用该方法在基准（x_0、y_0、α_{0-1}）下，进行点位径向位移的模拟分析，得到了如图2上部所示的点位径向位移模拟曲线。从图中可知，模拟曲线呈现一定的规律性。但为什么会呈现一定的规律性这一问题一直未得到满意的解决。

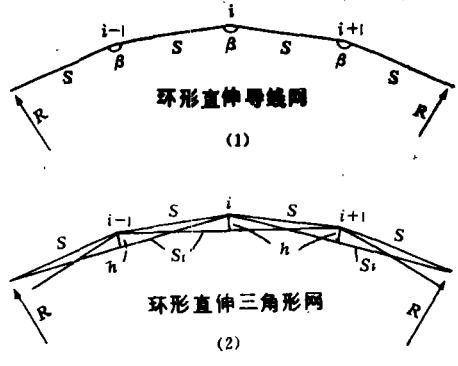


图 1

本文1987年7月收到.

*本文是研究生毕业论文的一部分，指导教师为陶本藻教授、杨仁副教授.

同时，对一个控制网，即使借助计算机进行大量的模拟试验，也是既费时，又费钱的事，且得到的结果还无完善的解释。为此，作者试图应用多元统计分析中的主分量（亦称主成分）分析理论，并结合模拟试验，以加速器环形控制网为例，对上述问题进行研究，所获结果是令人满意的，在此基础上，对如何解释模拟曲线的特定规律提出了自己的看法。

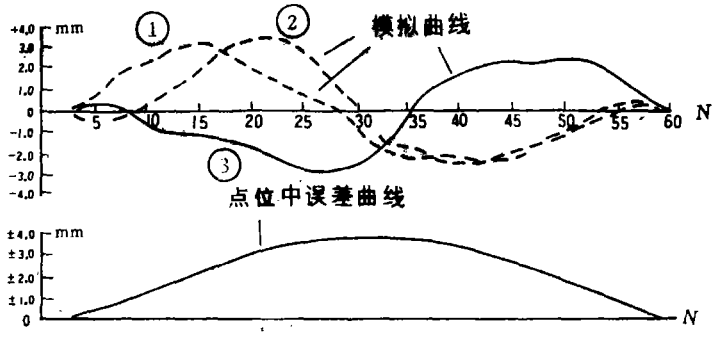


图 2

众所周知，主分量分析（Principal Component Analysis，简称 PCA）理论在测量控制网的灵敏度分析[4]、粗差检验、质量评定等方面已有应用。作者在此从完全不同的角度，应用它研究控制网中观测误差在待定参数及其函数上的分布规律。

2 主分量分析理论概述

设 X 为 t 维随机向量，其协方差阵为 $\sum_{XX}$（$\geqslant 0$）。现试图找到一组新变量 Z，使 Z 更能集中反映 X 的特征。

根据矩阵的谱分解理论[3]可知，$\sum_{XX}$ 可由它的特征值 λ_1、λ_2、……λ_t 及其相应的特征向量 u_1、u_2、……u_t 表示为

$$\sum_{XX} = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots\cdots + \lambda_t u_t u_t^T \tag{1}$$

其矩阵形式为

$$\sum_{XX} = U^T \wedge U \tag{2}$$

式中 $U^T = (u_1, u_2, \cdots, u_t)$，$\wedge = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_t)$，且

$$u_i^T u_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad (i,j = 1,2,\cdots,t) \tag{3}$$

不失一般性，设 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_t \geqslant 0$。若令

$$Z = UX \tag{4}$$

则 Z 即要找的一组新变量。

可以证明，对任意的向量 b，有

$$\max \mathrm{Var}(b^T x) = \lambda_1 = \lambda_{max}(\sum_{XX}) \tag{5}$$
$$b^T b = 1$$

当 $b=u_1$ 时

$$\mathrm{Var}(Z_1)=\lambda_1 \tag{6}$$

上式表明，参数向量 X 的一切线性组合中，对标准化的系数向量，Z_1 方差最大，相应的 u_1 的方向则代表了 X 在参数空间中变化最大的方向；继之，在与 u_1 不相关的任意线性组合中，Z_2 方差最大，相应的 u_2 是第二个变化最大的方向；依此类推。为此，多元统计分析中，依次称 $\sqrt{\lambda_1}u_1$，$\sqrt{\lambda_2}u_2$，$\cdots\sqrt{\lambda_t}u_t$ 为 $\sum_{XX}$ 的第一、第二、…第 t 主分量。称比值

$$\lambda_i/\mathrm{Tr}(\sum_{XX})\cdot 100\% \tag{7}$$

为第 i 主分量对 $\sum_{XX}$ 的贡献率。。称比值

$$(\lambda_1+\lambda_2+\cdots+\lambda_k)/\mathrm{Tr}(\sum_{XX})\cdot 100\% \qquad (k\leqslant t) \tag{8}$$

为前 k 个主分量的累计贡献率。贡献率是主分量分析中的主要数量指标之一。一般经验认为（〔9〕等）取累积贡献率为80%至85%的前 k 个主分量进行讨论即可，而严格的统计检验是相当复杂的。累积贡献率究竟多大为宜，应视具体问题而定。

如果对于某个 $k(\leqslant t)$，有 λ_k 之后的 $(t-k)$ 个特征值都显著地小，就认为随机向量 X 在参数空间中的随机变化（亦称变异[1]或变差[2]），近似地落在 k 维空间中。从几何观点看，一个 t 维椭球只在 k 个主轴方向上主轴有一定的长度，而在其它主轴方向上主轴长度可以忽略不计，因而考虑它们已无实际意义。这样，便可用前 k 个主分量来描述 X。

以上论述表明，一定存在一个 k 使 $\sum_{XX}$ 的前 k 个主分量包含 $\sum_{XX}$ 的主体部分，能够反映 X 随机变化的主要特征。作者正是以此为依据展开探讨性工作的。在实际工作中，如有可能，首先考虑 λ_{max} 及 $\sqrt{\lambda_{max}}u_1$，即第一主分量是方便的。此时 $k=1$，且

$$\sum_{XX}\approx\lambda_{max}u_1u_1^{\mathrm{T}} \tag{9}$$

当然，其效果取决于第一主分量的贡献率。本文是从选第一主分量着手的。

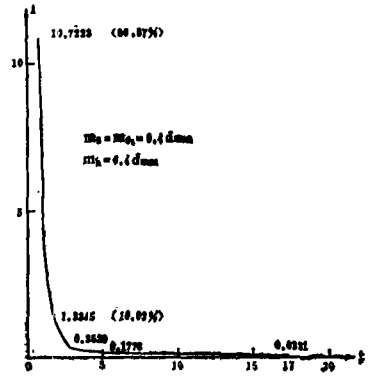
3 加速器环形网点位径向主分量分析计算公式

主分量分析在任何控制网的分析中均可进行，本文以加速器环形控制网为例。

3.1 环形控制网在基准（x_0，y_0，α_{0-1}）下 $\sum_{XX}$ 的特征值分布曲线

为有足够的理由解释为何选用第一主分量进行分析，先考察图1所示环形控制网在基准（x_0，y_0，α_{0-1}）下 $\sum_{XX}$ 的特征值分布情况。

以点数 $N=10$、半径 $R=50$ 米为例，分别计算两种方案 $\sum_{XX}$ 的非零特征值。它们的第一主分量贡献率分别为80.78%和80.57%，并绘出环形直伸三角形方案的特征值分布曲线（图3）。可见，该种情况下取第一主分量分析是可行的。



环形直伸三角网λ($\sum_{XX}$) 分布曲线(基准 (x_0、y_0、α_{0-1}))

图 3

3.2 加速器环形控制网点位径向主分量分析计算公式

若取 $X=(x_0,\ y_0,\ x_1,\ y_1,\cdots,x_{N-1},\ y_{N-1})^{\mathrm{T}}$，

$\sum_{XX}=\sigma_0^2 Q_{XX}$，则前述主分量分析的各个公式，均成为控制网点位坐标 x、y 的主分量分析公式。因此，$\sqrt{\lambda_{max}}u_1$ 便是相应的第一主分量。

环形网中，点位的径、切向坐标 R、T 是点位坐标 x、y 的函数：

$$\begin{pmatrix} R_i \\ T_i \end{pmatrix}=\begin{pmatrix} \cos\alpha_i & \sin\alpha_i \\ -\sin\alpha_i & \cos\alpha_i \end{pmatrix}\begin{pmatrix} x_i \\ y_i \end{pmatrix} \tag{10}$$

$i=0, 1, \cdots, N-1$，α_i 为点 i 处径向的方位角。因式（10）具有正交变换的关系，所以点位径、切向第一主分量可按下式

$$\begin{pmatrix} \xi_i \\ \eta_i \end{pmatrix}=\begin{pmatrix} \cos\alpha_i & \sin\alpha_i \\ -\sin\alpha_i & \cos\alpha_i \end{pmatrix}\begin{pmatrix} \sqrt{\lambda_{max}}\cdot u_{1x_i} \\ \sqrt{\lambda_{max}}\cdot u_{1y_i} \end{pmatrix} \tag{11}$$

直接求出。式中 ξ_i、η_i 分别表示点位径、切向第一主分量的元素，u_{1x_i}、u_{1y_i} 分别是u_1中 x_i、y_i 对应的元素。本文仅考虑径向部分，故取各点相应的 ξ_i，记为

$$\xi=(\xi_0、\xi_1、\cdots、\xi_{N-1})^T \tag{12}$$

ξ便是点位径向第一主分量。按点的序号，用ξ的各元素绘成曲线，便得到点位径向主分量曲线。显然，$-\xi$也是第一主分量。

4 环形控制网点位径向主分量曲线与点位径向位移模拟曲线的比较

本文的目的就是研究大量模拟曲线所具有的统计规律是否同主分量曲线的分布规律相一致。为此，对图 1 所示的两种环形控制网方案，取 $R=50$米、$N=30$、$m_\beta=1$秒、$m_S=m_{S_L}=m_h=40$微米，进行分析如下：

4.1 点位径向主分量曲线

经计算给出两种方案在基准(x_0, y_0, α_{0-1})下的点位径向第一主分量曲线（图4、5）。相应的第一主分量贡献率为：环形直伸导线为81.4%，环形直伸三角形为84.8%，都是很大的。同时也绘出了点位中误差曲线和点位径向中误差曲线（图4、5）。

4.2 点位径向位移模拟曲线

作者按数学模型扭曲法，分别对环形直伸导线和环形直伸三角形各进行了六十次点位径向位移模拟试验。现将作法简述如下：

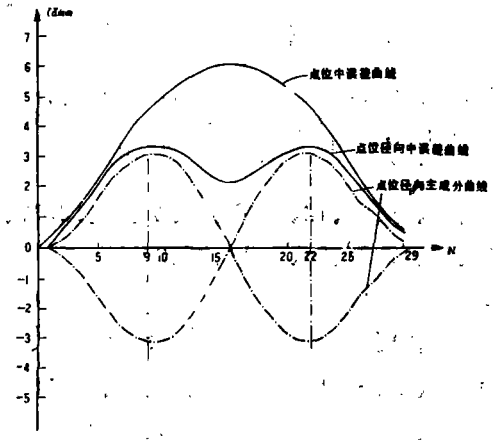
若记观测值的真值为 $\widetilde{L}$、未知数的真值为$\widetilde{X}$、误差方程式系数为 B，则有

$$\widetilde{L}=B\widetilde{X} \tag{13}$$

现利用一组正态随机数 ∇l（即 $\nabla l\sim N(0, \sigma_0^2 Q_{LL})$），作为模拟的偶然误差来扭曲 $\widetilde{L}$，此时，$\widetilde{X}$ 将产生相应的变化量 ∇x：

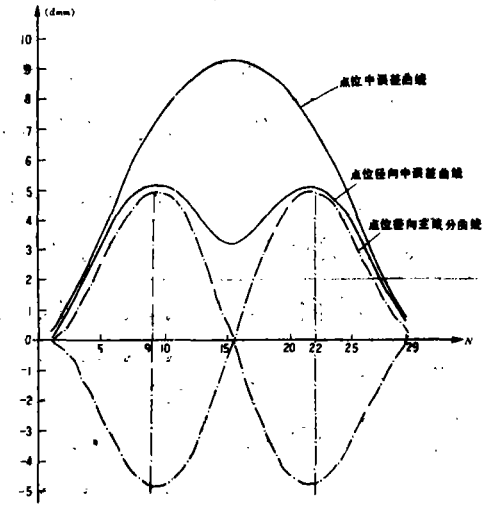
$$\nabla x=Q_{XX}B^TP\nabla l \quad (P=Q_{LL}^{-1}) \tag{14}$$

再利用式(10)的关系及式(12)类似的方法，便可得到一组点位径向位移模拟值 ∇R



环形直伸导线网

图 4



环形直伸三角形网

图 5

$$\nabla R = (\nabla R_0,\ \nabla R_1,\ \cdots,\ \nabla R_{N-1})^T \tag{15}$$

将 ∇R 的各个分量按点绘成曲线，就是点位径向位移模拟曲线。其中正态随机数 ∇l 是根据中心极限定理，借助计算机内部的随机函数编程获得的，有关的理论及生成 ∇l 时的统计检验内容见〔8〕。

详细的模拟结果略，仅给少量示例（图6—1、图6—2和图6—3等）。

4.3 比较和分析

位移模拟曲线与主分量曲线按基本符合、一般和不符合三种情形进行比较。其划分的标准主要有如下几点：

1、小于主分量曲线峰值 1/3 的位移模拟曲线峰点不计。

2、若位移模拟曲线同主分量曲线（图 4 和图 5）形状相同，且中部的零点位置以及正、负峰值点位置，同主分量曲线相比偏离不超过 ± 3 个点时，则认为二者基本符合。

3、若位移模拟曲线同主分量曲线相比，除中部的零点位置以及正负峰值点位置偏离超过 ± 3 个点外，其它同“2”，则认为二者的符合情况一般。

4、若位移模拟曲线同主分量曲线相比，形状不一致(如有多个峰值点、形状杂乱等)，则认为二者不符合。

比较结果列于表 1。为明了起见，特给出三种情形的示例(图6—1～图6—3)，供参考。比较结果是令人满意的，环形直伸导线和环形直伸三角形两种方案的点位径向位移模拟曲线同各自相应的点位径向主分量曲线相比，不符合率都在16.7%左右。另外，从二者的曲线峰值来看，位移模拟曲线的峰值基本小于主分量曲线峰值的 3 倍（即取 $3\sigma_0$），比较结果列于表 2。

这表明，位移模拟曲线所具有的统计规律与相对应的主分量曲线的分布规律相比，是基本吻合的，即主分量分析可以揭示偶然误差在待定参数及其函数上的分布规律，集中地体现

位移模拟曲线的固有特征。

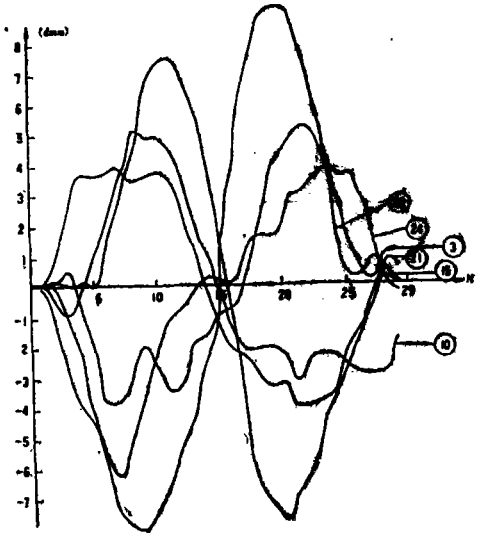
表 1　径向位移模拟曲线与径向主分量曲线形状比较

网名			环形直伸导线		环形直伸三角形	
基准			(x_0, y_0, α_{0-1})		(x_0, y_0, α_{0-1})	$(y_0, x_0, \alpha_{0-15})$
λ_{max} 的贡献率			81.4%		84.8%	57.0%
第一组试验（30次）	基本符合	曲线序号	0、2、3、5、7、8、9、10、11、13、15、16、18、20、21、22、26、29、		3、4、5、8、10、12、13、15、16、17、21、24、25、28	0、2、7、11、12、17、19、27、29
		百分比	60%		46.7%	30%
	一般	曲线序号	1、4、12、14、17、24、25、28		0、1、2、7、9、11、14、18、19、20、23、27、29	6、8、10、14、15、16、20、22、26
		百分比	26.7%		43.3%	30.0%
	不符合	曲线序号	6、19、23、27		6、22、26	1、3、4、5、9、13、18、21、23、24、25、28
		百分比	13.3%		10%	40%
第二组试验（30次）	基本符合	曲线序号	3、4、、6、8、9、10、11、12、13、14、15、20、23、27、29		0、2、3、5、6、8、12、13、14、19、21、22、25、27、29	
		百分比	50%		50%	
	一般	曲线序号	0、2、5、7、16、17、22、26、28		1、7、9、15、17、20、24、26	
		百分比	30%		26.7%	
	不符合	曲线序号	1、18、19、21、24、25		4、10、11、16、18、23、28	
		百分比	20%		23.3%	
总计	基本符合		55%	83.3%	48.3%	30.0%
	一般		28.3%		35.0%	83.3% 30.0% 60%
	不符合		16.7%		16.7%	40.0%

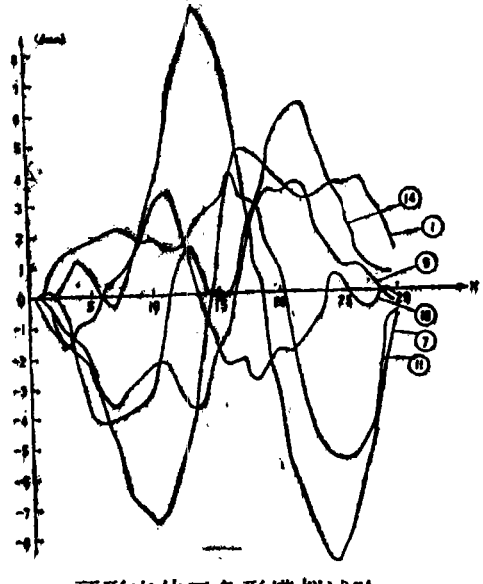
表2　　峰值比较

网名		环形直伸导线	环形直伸三角形	
基准		(x_0, y_0, α_{0-1})	(x_0, y_0, α_{0-1})	$(x_0, y_0, \alpha_{0-15})$
模拟曲线峰值		百分比	百分比	百分比
第一组试验	$\leqslant \lvert\xi_i\rvert^*_{max}$	26.7%	43.3%	20%
	$\leqslant 2\lvert\xi_i\rvert_{max}$	86.7%	93.3%	80%
	$\leqslant 3\lvert\xi_i\rvert_{max}$	100%	96.7%	100%
	$>3\lvert\xi_i\rvert_{max}$	0	3.3%	0
第二组试验	$\leqslant \lvert\xi_i\rvert_{max}$	33.3%	46.7%	
	$\leqslant 2\lvert\xi_i\rvert_{max}$	93.3%	90.0%	
	$\leqslant 3\lvert\xi_i\rvert_{max}$	100%	100%	
	$>3\lvert\xi_i\rvert_{max}$	0	0	

*　表示取一倍 σ_0 时主分量曲线的峰值

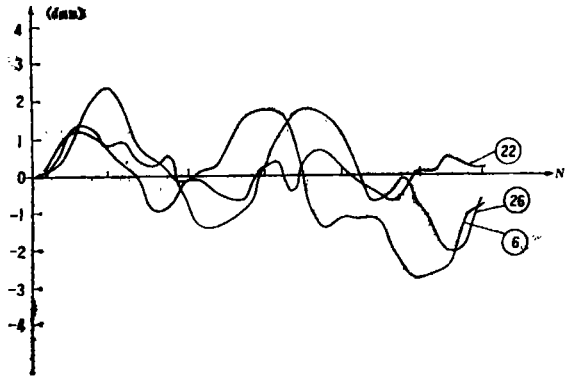


环形直伸三角形模拟试验
第1组示例：基本符合
图 6—1



环形直伸三角形模拟试验
第1组示例：一般
图 6—2

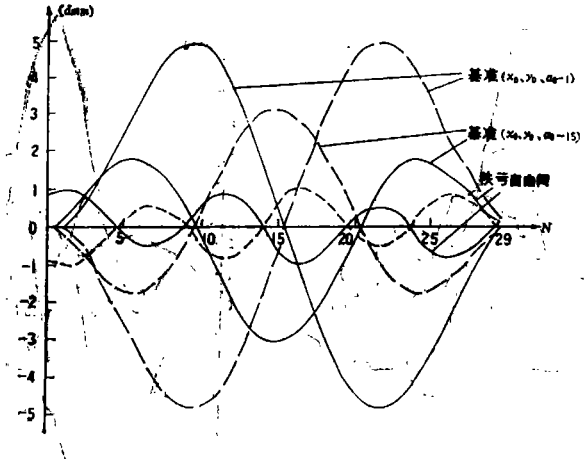
是否在任何基准下，第一主分量都具有这样的特性呢？作者进一步将环形直伸三角形网的第一组30次试验，分别转换至基准 $(x_0、y_0、\alpha_{0-15})$ 及秩亏自由网下进行分析比较，看出，三种基准下点位径向主分量曲线变化很大（图7）；各 λ_{max} 的贡献率变化也很大，在基准 $(x_0、y_0、\alpha_{0-15})$ 下为57.0%，而秩亏自由网时已降为 40.5%。在基准 $(x_0、y_0、\alpha_{0-15})$ 下，两种曲线的比较结果也列于表1和表2，此时符合率（80%）同 λ_{max} 的贡献率



环形直伸三角形模拟试验第 1 组示例：不符合

图 6 — 3

（57%）相比，也是令人满意的，即此时的主分量曲线仍能揭示位移模拟曲线的规律。但对秩亏自由网平差的结果而言，第一主分量曲线已不再具有这种特性，或许应考察更多的主分量进行更深入的研究。



环形直伸三角形网点位径向主分量曲线

图 7

5 加速器环形控制网点位径向位移模拟曲线的意义

通过上述分析，作者对加速器环形控制网点位径向位移模拟曲线的含义及其使用中的某些问题，提出如下几点看法：

1、点位径向主分量曲线揭示了点位径向位移模拟曲线具有特定分布规律的内在原因，即这种规律性取决于$\sum_{xx}$，亦即网形、观测精度和平差基准。这为合理地解释位移模拟曲线提供了严密的理论依据。

2、主分量曲线和位移模拟曲线能同时反映大小、正负和方向等因素，其优点是中误差

曲线所无法比拟的。因此，它们对于揭示控制网中偶然误差的分布规律，更具有实用性和严密性。

3、将中误差曲线同位移模拟曲线直接进行比较是不恰当的。因为，任何中误差曲线都只能是一种包络线，实用中应充分注意到这一点（图4、5）。

4、文献〔6〕中关于加速器环形控制网点位径向位移模拟曲线（图2）的下述解释，作者认为是不恰当的："在所有的三组中，点的最大（径向）位移都不发生在导线中部通常认为点位可靠性最弱（原意指精度最弱——作者注）的地方"。得出这一结论的原因是，文献中将三条点位径向位移模拟曲线同点位中误差曲线（图2下部）进行比较。如果将其同点位径向中误差曲线进行比较（图4、5），两种曲线的幅度是基本一致的，但在其它方面也是不同的。上述不恰当的结论，曾被国内许多文献所引用，应加以更正为宜。

6 結　论

作者利用主分量分析理论研究控制网中偶然误差的分布规律，试验结果表明，当第一主分量贡献率较大时（文中实例为80%～85%及55%左右），有如下结果：

1、主分量分析与数学模型扭曲法具有一致性，即主分量曲线同位移模拟曲线分布规律是相同的。

2、利用主分量分析研究观测偶然误差在控制网中待定参数及其函数上的分布规律，既简单，又有效，它可以避免大量的模拟试验。因此，主分量分析是代替按数学模型扭曲法进行模拟分析的一种有效方法。建议实际工作中采用。

3、对第一主分量贡献率小的情况，还需考虑多个主分量进行更深入的研究，以使该方法不断完善。

参 考 文 献

〔1〕 M.Kendall著，中国科学院计算中心概率统计组译，多元分析，科学出版社，1983

〔2〕 王松桂，林春土，主成分的最优性质，科学通报，No.8，1984

〔3〕 何旭初，广义逆矩阵的基本理论和计算方法，上海科技出版社，1985

〔4〕 张正禄，变形监测网的灵敏度分析，武汉测绘科技大学学报，No.3，1986

〔5〕 郑国忠，用数学模型扭曲法设计精密工程测量控制网，工程勘察，No.3，1982

〔6〕 Большаков В.Д., Горбенко О.И., Климов О.Д. Высокоточные геодезические измерения для строительства и монтажа Большого Серпуховского ускорителя. М., Недра, 1968

〔7〕 В.Д.鲍利沙科夫著，孔祥元译，精密工程测量的方法和仪器，测绘出版社，1981

〔8〕 郑肇葆，正态分布伪随机数据的产生和检验，武汉测绘学院学报，No.1，1980

〔9〕 唐守正编著，多元统计分析方法，中国林业出版社，1986

Experimental Research on the Consistence between PCA Method and Distorting Method of Mathematical Model

Wang Jianguo

Abstract

In this paper, the consistence between PCA method and distorting method of the mathematical model in the study of survey networks is researched. As a practical example, the simulating experiments of an accelerator ring-shaped network are made. The results show that both methods are consistent under some conditions. It not only reveals the reason for specific regularity existing in the simulating curves, but also find an effective way to avoid a lot of simulating experiments. Meanwhile some author's views are given.

[Key words] principal component analysis (PCA) ; distorting method of mathematical model; ring-shaped network; radial principal component curve of points; radial displacement simulating curve of points