

参赛密码 _____

(由组委会填写)



“华为杯”第十三届全国研究生 数学建模竞赛

题 目 基于无线通信基站的预估-校正-自适应定位模型

摘 要：

本文主要对基于无线通信基站的室内三维定位问题展开研究，在研究了 TOA 定位原理、多基站联合定位算法、非视距传播下的 TOA 定位原理、信道传输损耗模型、极大似然估计方法的基础上，提出了一种新的定位模型。

本文在分析相关文献的基础上，首先对测得的手持终端到基站的 TOA 数据算得的距离进行了预估，给出了距离误差因子 a ；其次我们考虑信号传播过程中的信号干扰比值、多径传播、终端与基站时钟不同步等因素，利用 Shadowing 和 hata 两个信道传播路径损耗模型，对距离进行自适应的修正，得到较为精确的距离 d ；最后改进后的极大似然估计方法估计出精度较高的终端位置。由于本文使用的信道传输损耗模型具有自适应识别 LOS 和 NLOS 传播环境的特性，因此本文将新模型称为“预估-校正-自适应定位模型”。

对于问题一，本文利用预估-校正-自适应定位模型，求解给出了 10 组 LOS 或 NLOS 传播环境下手持终端的三维坐标。

对于问题二给出的 10 个场景，本文从每个场景的全部基站中分别随机选取 $i(i=4, \dots, 10)$ 个基站，进行了 20 组定位实验并得到手持终端坐标的均方根误差 RMSE 值。当 $n=7$ 时，模拟出的终端三维精度误差控制在 $5m$ 。当 $n=8, 9, 10$ 时，

较之 $n=7$ 时模拟出的终端坐标 $RMSE$ 值相差无几，即精度无明显提升。因此本文选择 7 或 8 个终端数，来实现近似最优的三维定位精度。

对于问题三，本文利用上述模型，选取 7 个基站作为参考基站实现了对移动过程中的终端二维定位，拟合并计算出终端的移动轨迹方程。

对于问题四，本文通过给定的 TOA 数据，利用模型筛选出小于 200 米的有效数据。统计各个终端可用的 TOA 数据的数目（即连接数），寻找出了可以被基站定位的连接数大于等于 4 的所有终端。本题中的精度采用均方根误差的倒数。本题仍采用预估-校正-自适应模型，拟合出连接度数与定位精度之间的关系式。

关键词：预估-校正-自适应，TOA 定位模型，路径损耗模型，极大似然估计

目录

一、	问题重述.....	1
二、	模型假设.....	2
三、	相关定义与符号说明.....	2
	3.1 相关定义.....	2
	3.2 符号说明.....	4
四、	TOA 定位原理.....	4
	4.1 理想情况下的 TOA 定位原理	4
	4.2 非视距传播下的 TOA 定位原理	6
五、	预估-校正-自适应定位模型的建立	7
	5.1 多基站联合定位算法.....	7
	5.2 信道传输损耗模型.....	8
	5.3 极大似然估计.....	9
	5.4 预估-校正-自适应定位模型.....	11
六、	问题分析与求解.....	12
	6.1 问题一的分析与求解.....	12
	6.2 问题二的分析与求解.....	15
	6.3 问题三的分析与求解.....	16
	6.4 问题四的分析与求解.....	19
七、	模型的不足与展望.....	21
	参考文献.....	22

一、 问题重述

随着无线通信网络和移动互联网的蓬勃发展，提供基于地理位置信息的服务（Location Based Service，简称 LBS）已经成为最具市场前景和发展潜力的业务之一。从传统的 GPS 导航，到大众点评、微信等基于地理位置的消费信息服务和社交软件，实现其功能的基础就是要通过手机、导航仪等终端设备收发信号，来获得距离、角度等测量信息，并利用定位算法将这些测量信息转换成坐标信息。

基于无线移动通信网络的定位是以获取用户手持终端（包括手机或者平板等设备）的位置为目标。而达成这一目标的手段是通过测量无线电信号的强度、传播时间、到达角等物理指标，并将其转化成终端与基站之间的距离、角度等信息，最终利用定位算法将距离、角度等信息转化成终端的坐标信息。

手持终端设备如何基于基站的测量信息，计算或确定终端在三维空间中的位置坐标，也就是三维定位问题，被认为是现代商用通信网络中对于定位系统真正具有技术难度的挑战。而高精度三维定位也预期能为客户提供更大的价值，在智能仓储、智能工厂、固定资产追踪等对于三维坐标信息敏感的垂直行业，以及传统运营商感兴趣的商场、办公楼中基于位置信息的室内导航、人群流量分析，以及基于精确三维地理位置信息的业务推送等服务提供基础性技术。

从技术角度来看，现代商用通信网络对于三维定位的需求，是使用尽可能少的基站完成对终端设备的定位、算法收敛速度快、对于干扰和噪声具有鲁棒性等优点。

在本题中，需要解决如下四个方面的问题：

1. 给定 10 组 LOS 或 NLOS 传播环境下从手持终端到基站的 TOA 测量数据和所有基站的三维坐标（对应附录中编号为 case001_input.txt 到 case010_input.txt 的文件），请根据这些测量数据计算出终端的三维坐标。（请给出详细的建模分析，建模过程中建议考虑测量模型、误差分析等内容。）
2. 给定 10 组 TOA 测量数据和所有基站的三维坐标（对应附录中编号为 case011_input.txt 到 case020_input.txt 的文件），请设计算法，使用尽可能少的基站数目，实现近似最优的三维定位精度。
3. 给定 5 组对处于移动过程中的终端采集到的 TOA 数据（对应附录中编号为 case021_input.txt 到 case025_input.txt 的文件），请设计算法计算出终端的运动轨迹。（此时，编号为 case021_input.txt 到 case025_input.txt 的文件中，只记录一个终端的 TOA 数据，并且是这一个终端在运动轨迹中多个位置上的 TOA 数据。）
4. 在前述 3 问中，都是假设给定区域内终端到每一个基站的距离都是可知的，但事实上，基站的通信半径是有限的，因此，只有在基站通信半径覆盖范围内的终端才有可能测到自身到基站的距离。而一个终端只有获得它与足够数目的基站之间的距离测量值，才能完成定位。假设每个基站的通信半径为 200 米（超过范围虽然有测量数据，但无效）。请根据给定的 5 组测量信息数据集（对应附录中编号为 case026_input.txt 到 case030_input.txt 的文件），设计算法寻找出可以被基站定位的所有终端。进一步，回答如下问题：每一个场景中（对应着 case026_input.txt 到 case030_input.txt 五个文件中的一个），定义终端的平均“连接度数” λ 为

$$\lambda = \frac{\text{所有可以被定位终端到基站之间的连接数}}{\text{终端数}}$$

请建立模型分析连接度数 λ 与定位精度之间的关系。

二、 模型假设

1. 由于同一材料或电子器件的时钟是同步的，因此可认为手持终端之间的时钟是同步的，基站之间的时钟是同步的，但手持终端与基站之间存在时差；
2. 基站测量时间时存在测量上的误差，且测量误差服从均值为零的正态分布；
3. 无线电信号的传播速度统一取 $3 \times 10^8 \text{ m/s}$ ；
4. 无线电信号在手持终端发送和基站接收过程中不存在时间延迟，即无线电信号只在传播过程中由于环境的影响产生延迟；
5. 假定题目中给出的每一组数据都是在相似环境中测出的，因此可以认为同一组的传播环境相同。

三、 相关定义与符号说明

3.1 相关定义

1. 视距（LOS）与非视距（NLOS）传播

无线电信号在大气中从 A 点向 B 点传播时，如果传播过程中存在一个没有遮挡的直达路径，那么，这种传播环境被称为视距传播环境（Line Of Sight propagation，简称 LOS）。这种传播环境如图 1 中的左图所示。如果在传播过程中，由于建筑物或树木的遮挡、反射、折射等物理现象，使得从 A 点向 B 点之间存在多条无线电信号的传播路径，这种环境被称为非视距传播环境（Non-Line Of Sight，简称 NLOS）。

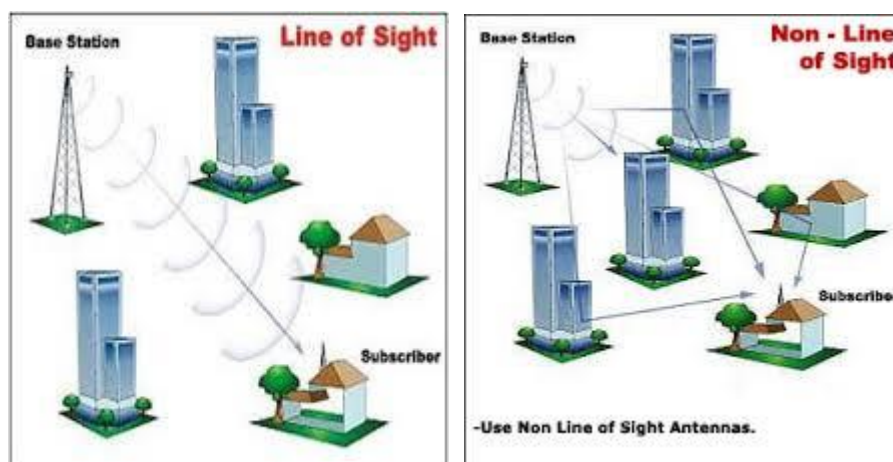


图 1 LOS 径与非 LOS 径示意图

2. 到达时间（TOA）测量

当无线电信号在基站与用户手持终端之间互相传播时，就可以计算基站与手持终端之间的距离，一种常用的测量方式是记录无线电信号从手持终端发出，直到基站接收到信号为止的无线电信号传播时间，将时间乘以无线电信号的传播速度，即得到基站与终端之间沿某条路径的距离。其中，信号在基站与终端之间的传播时间，被称为无线电信号的到达时间（Time Of Arrival，简称 TOA），如图 2 所示。

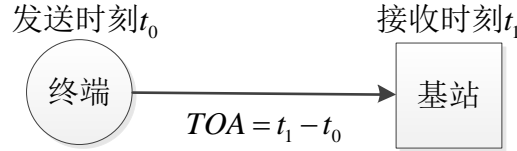


图 2 到达时间（TOA）测量示意图

3. 信号干扰比值（SINR）

$$SINR = \frac{\text{有用信号强度}}{\text{干扰信号强度} + \text{噪声信号强度}}$$

4. 多径传播（multi-path propagation）

室内环境下，由于反射频繁发生，会形成无线电波的多径传播，从而导致虽然距离很近接收到的信号强度却波动剧烈。

5. 连接度数 λ

$$\lambda = \frac{\text{所有可以被定位终端到基站之间的连接数}}{\text{终端数}}$$

6. 拟合误差

假设某一终端在时刻 i 的位置为 (X_i, Y_i) ，则该终端所有时刻的位置组成的向量可记作 (X, Y) ，由 (X, Y) 拟合出的方程轨迹满足 $\bar{Y} = f(X)$ ，则拟合误差可以用 $Err = \|\bar{Y} - Y\|_{\infty}$ 表示。

7. 均方根误差（RMSE）

均方根误差，它是观测值与真值偏差的平方和观测次数 n 比值的平方根，在实际测量中，观测次数 n 总是有限的，真值只能用最可信赖（最佳）值来代替。方根误差对一组测量中的特大或特小误差反映非常敏感，所以，均方根误差能够很好地反映出测量的精密度。当对某一量进行多次测量时，取这一测量列误差的均方根误差（误差平方的算术平均值再开方），亦称标准偏差。**RMSE** 反映了测量数据偏离真实值的程度，**RMSE** 越小，表示测量精度越高，因此可用 **RMSE** 作为评定这一测量过程精度的标准，其计算公式为

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{o,i} - X_m)^2}{n}}.$$

3.2 符号说明

表 1 符号说明

符号	意义
BS	基站
A, B, C, D	手持终端的定位位置
A', B', C', D'	手持终端的实际位置
x, y, z	终端或基站三维坐标
r	TOA 测量距离
d	终端与基站的实际距离
a	距离误差因子

其他符号及意义将在文章中给出。

四、TOA 定位原理

4.1 理想情况下的 TOA 定位原理

在不存在测量误差的理想情况下，假设 (x_0, y_0, z_0) 为某个基站的坐标，终端的位置为 (x, y, z) ， TOA 定位方法[1]的数学表达式可以表示为：

$$d^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$$

其中 r 为所测的距离， d 是终端和基站之间的实际距离(在理想情况下， $d = r$)。单个基站 TOA 定位模型图示如下：

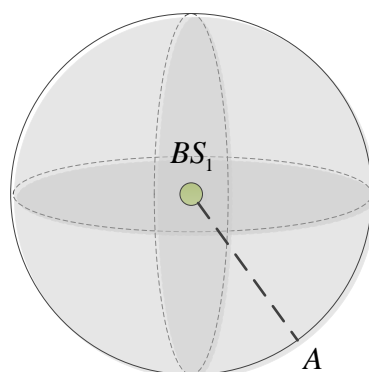


图 3 单个基站 TOA 定位模型

理想情况下，根据所给的从手持终端到某一个基站的 TOA 数据，便能确定终端在以基站 (x_0, y_0, z_0) 为圆心， r 为半径的球上。

在存在两个基站 (x_0, y_0, z_0) 和 (x_1, y_1, z_1) 时，可得到类似的方程组，即

$$\begin{cases} d_0^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r_0^2 \\ d_1^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = r_1^2 \end{cases}$$

其定位原理如下图：

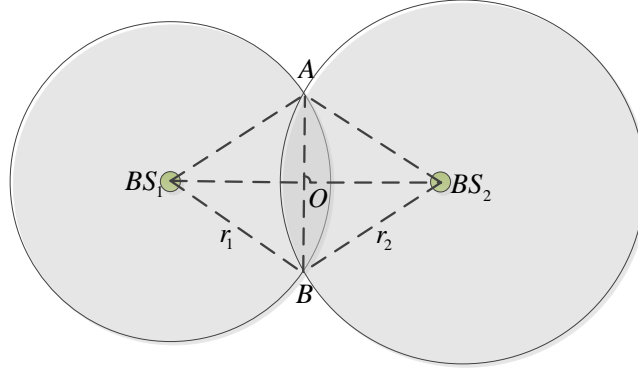


图 4 两个基站定位模型（平面平面图）

理想情况下，两个基站定位，只能确定终端位于两个球相交的圆上，即以 O 为圆心、以 AB 为直径的圆上，还是无法准确地获得终端的位置。

在存在三个基站时，基站位置坐标分别为 (x_0, y_0, z_0) 、 (x_1, y_1, z_1) 和 (x_2, y_2, z_2) ，可得

$$\begin{cases} d_0^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r_0^2 \\ d_1^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = r_1^2 \\ d_2^2 = (x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 = r_2^2 \end{cases}$$

定位原理如下图

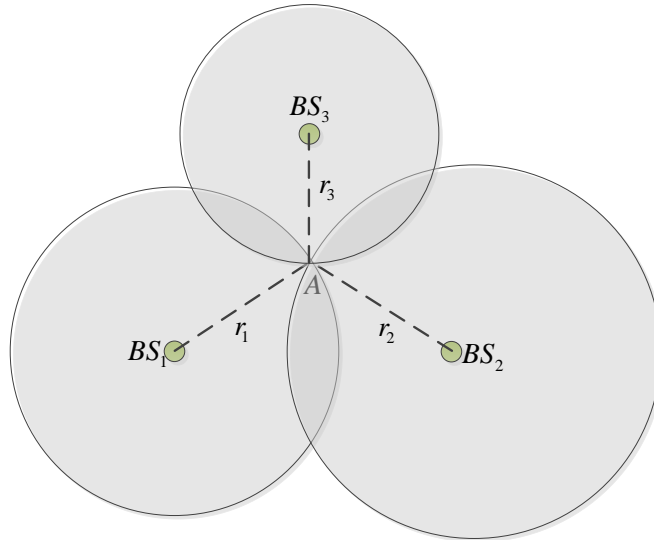


图 5 三个基站定位模型（平面投影图）

此方程组只能确定终端位于三个球相交的点上。由于相交的点不止一个（假设球两两相交），故仍然无法确定终端的位置。

在存在四个基站时，基站位置坐标分别为 (x_0, y_0, z_0) 、 (x_1, y_1, z_1) 、 (x_2, y_2, z_2) 和 (x_3, y_3, z_3) ，则可得

$$\begin{cases} d_0^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r_0^2 \\ d_1^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = r_1^2 \\ d_2^2 = (x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 = r_2^2 \\ d_3^2 = (x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2 = r_3^2 \end{cases}$$

当该方程组的解存在时，解是唯一的。因此在理想情况下，*TOA*定位需要有四个基站，且终端的位置在四个球的交点上。

4.2 非视距传播下的*TOA*定位原理

存在NLOS传播时，NLOS误差在*TOA*定位方法中是一个正的附加延时超量，所以测得的终端到基站的距离就会大于终端到基站的实际距离。

对于单个基站来说，定位原理如图所示。

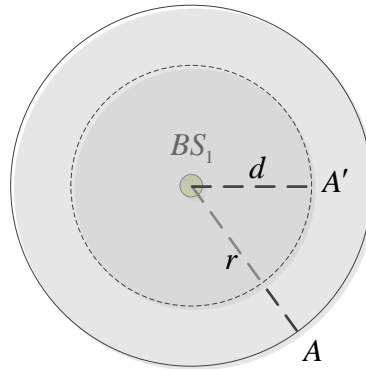


图6 非视距传播下的单基站定位模型（平面投影图）

可以得到下面的式子：

$$d^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = ar^2$$

其中 r 为所测的距离， d 是终端与基站之间的实际距离， a 表示距离误差因子。由于 $d \leq r$ ，故 $0 < a \leq 1$ 。当 $a = 1$ 时， $d = r$ ，即该式表示的是无线电信号在视距传播的情况。

对于存在两个基站 (x_0, y_0, z_0) 和 (x_1, y_1, z_1) 时，定位原理图如图7所示。

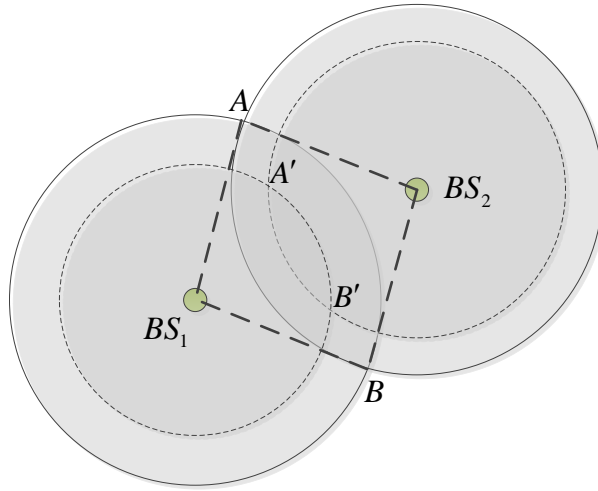


图 7 非视距传播下的双基站定位模型（平面投影图）

由于 NLOS 误差的存在，测得的移动台和两个基站之间距离为 r_0 和 r_1 ，它们的值大于或等于移动台与两个基站之间得实际距离 d_0 和 d_1 ，所以只能确定移动台的位置在图 7 中两个圆的公共部分。它的范围较大，定位信息又较少，难以确定移动台的位置。

与理想情况下的四基站 TOA 定位模型相比，NLOS 传播下的四基站 TOA 模型中，四个球没有公共交点，而是相交于一个范围，终端在这个范围里，仍然不能确定它的准确位置坐标。但它比三基站定位模型中确定的定位范围要小，而且定位信息较多，可以通过一些改进的方法对移动台的位置进行定位估计。

在非视距传播下，上面的定位方法只用到了少数基站，而在实际情况下，接收到信息的基站往往是所有的基站，有些基站可能距离终端很远，有的则很近，位置分布很不均匀，由于 TOA 测量误差的正均值通常具有随着终端和基站之间的距离呈现线性增加的趋势，造成它们的 NLOS 的差别非常大，这样基站的几何位置关系对最后的定位估计产生很大的影响。文献[1]提出选择基站定位算法，该算法通过对基站的选择以提高定位精度，但是由于引入基站选择算法导致算法复杂，定位不方便。文献[1]提出了多个基站联合定位的算法。

五、预估-校正-自适应定位模型的建立

5.1 多基站联合定位算法

假定终端向 n 个基站发送了信息，将基站测量的信息以四个基站作组合，可以得到 C_n^4 种组合，有多少种组合，就能使用基站定位算法得到多少个终端的位置初估计，最后通过对坐标值求平均值的方法进行优化。假设共有 m 种组合，则有 m 个初估计值，设其坐标分别为 (x_j, y_j, z_j) ， $j = 1, \dots, m$ 。设需要定位的终端的坐标位置为 (x_0, y_0, z_0) ，于是有，

$$\begin{cases} x_0 = \sum_{j=1}^m x_j / m \\ y_0 = \sum_{j=1}^m y_j / m \\ z_0 = \sum_{j=1}^m z_j / m \end{cases}$$

则 (x_0, y_0, z_0) 就是所求的终端位置，误差可以用 (x_j, y_j, z_j) 之间的均方根误差 *RMSE* 来近似。

算法的定位步骤如下：

- 1) 对所有的基站进行组合，四基站作为一个组合；
- 2) 将每个组合采用四基站 *TOA* 定位算法求出终端坐标的初估计；
- 3) 对初估计坐标值用求均值的方法进行优化，最终得到终端的位置坐标。

该算法存在多个基站信息的状况下，不对基站进行选择，不引入其他复杂的算法，就减小了由于不平衡的位置关系而引入的部分位置误差，保障了算法的稳健性。

5.2 信道传输损耗模型

考虑到从手持终端发射信号，基站接收信号过程中，信号由于 *NLOS* (非视距) 传播，折射以及绕射传播，周边环境的影响等因素造成了信号强度在传播过程中的路径损耗与多径传播。我们使用以下的信道传输损耗模型，即 *Shadowing* 模型[2]，来衡量非视距传播的影响：

$$PL(d) = PL_0 + 10n \lg\left(\frac{d}{d_0}\right) + S \quad (1)$$

其中， $PL(d)$ 为接收机与发射机距离为 d 时的路径传输损耗（单位：*dB*）， PL_0 为参考距离下的传播损耗， d_0 为参考距离，通常取值为 1 米， n 为路径损耗指数， S 为均值为 0 的高斯噪声。本题中我们将终端与基站由于时钟引起的不同步引起的 200ns 的误差定义为 S 。由于没有提供移动终端的发射强度以及基站的接收强度，因此我们从移动终端的发射频率出发，参考资料并结合本题的信息，构建奥村-哈塔 (*hata*) 路径损耗模型：

路径损耗可以表示为如下参数的函数：(a) 载波频率 f_c ；(b) 移动终端高度 h_{re} ；(c) 基站高度 h_{te} ；(d) 基站和移动终端之间的距离。

$$PL(d) = \begin{cases} A + B \lg(d) & (\text{市区}) \\ A + B \lg(d) - C & (\text{市郊}) \\ A + B \lg(d) - D & (\text{开阔地区}) \end{cases} \quad (2)$$

$$\text{其中, } \begin{cases} A = 69.555 + 26.16 \lg(f_c) - 13.82 \lg(h_{te}) - a(h_{re}) \\ B = 44.9 - 6.55 \lg(h_{te}) \\ C = 5.4 + 2[\lg(f_c / 28)]^2 \\ D = 40.94 + 4.78[\lg(f_c)]^2 - 18.33 \lg(f_c) \end{cases}$$

上式中 $a(h_{re})$ 为移动台天线高度的校正因子，有：

$$a(h_{re}) = \begin{cases} 8.29[\lg(1.54h_{re})]^2 - 1.1 & f_c \leq 300\text{MHz} \\ 3.2[\lg(11.75h_{re})]^2 - 4.97 & f_c \geq 300\text{MHz} \end{cases}$$

查阅资料，了解到终端上行（移动台发，基站收）的载波频率在 $899 \sim 915\text{MHz}$ 之间，本题取 $f_c = 905\text{MHz}$ ，相应的 $a(h_{re}) = 3.2[\lg(11.75h_{re})]^2 - 4.97$ 。

在初估移动终端位置的过程中，求得的距离误差因子 a 均较小，终端信号在传播过程均受到了 NLOS 传播的影响，因此推理传播所处的环境是市区，相应的 $PL(d) = A + B\lg(d)$ 。根据所掌握的资料与数据，联立 hata 模型和 Shadowing 模型，求得修正后的终端与基站距离。

在计算过程中，结合 Shadowing 模型中的路径损耗模型 n ，本文将信号干扰比值（SINR）引入了 Shadowing 模型：

$$PL(d) = PL_0 + 10\text{SINR}\lg\left(\frac{d}{d_0}\right) + S$$

即当信号传播过程中，如果 LOS 传播占主导，基站接收到的有用信号将远强于干扰信号与噪声信号，但是由于周边环境的原因，干扰信号不可能完全消失，我们取 SINR 极大值为 10。相反的，如果 NLOS 传播占主导，基站接收到的有用信号几乎与干扰信号一样，很大影响了基站测量的时间，我们取 SINR 极小值为 1。

为了能够实现自适应的选择与场景匹配的模式算法，并且能够根据测量数据迅速对终端进行准确定位，在联立 Shadowing 模型与 hata 模型计算移动终端与基站之间距离时，我们结合上一步估算出的距离误差因子 a ，计算出初估距离为 $\sqrt{a} \times d$ 。利用二分法自适应的计算 SINR 的值，即当 a 接近 0 时，NLOS 传播占主导，根据初估距离与二分法自适应计算出的 SINR 值接近 10。相反的，当 a 接近 1 时，LOS 传播占主导，自适应计算出的 SINR 值接近 1。因此，我们通过联立 Shadowing 模型和 hata 模型，引入路径损耗、LOS 与 NLOS 传播、时钟不同步等因素后，计算出根据传播环境与场景自适应匹配的修正后的移动终端与基站距离 d 。

5.3 极大似然估计

设三维空间的 n 个参考节点即基站的坐标为 (x_1, y_1, z_1) 、 (x_2, y_2, z_2) 、 (x_3, y_3, z_3) 、 \dots 、 (x_n, y_n, z_n) ，移动终端 X 的坐标为 (x, y, z) ，点 X 到第 1 个、第 2 个、 \dots 、第 n 个参考节点的距离分别是 d_1 、 d_2 、 \dots 、 d_n ，则根据空间两点间距离公式有方程组：

$$\begin{cases} (x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2 = d_1^2 \\ (x_2 - x)^2 + (y_2 - y)^2 + (z_2 - z)^2 = d_2^2 \\ \dots\dots\dots \\ (x_n - x)^2 + (y_n - y)^2 + (z_n - z)^2 = d_n^2 \end{cases}$$

上式中，从第一个方程开始分别减去最后一个方程并移项，可得非齐次线性

方程组：

$$\begin{cases} 2(x_1 - x_n)x + 2(y_1 - y_n)y + 2(z_1 - z_n)z = (x_1^2 + y_1^2 + z_1^2) - (x_n^2 + y_n^2 + z_n^2) - (d_1^2 - d_n^2) \\ 2(x_2 - x_n)x + 2(y_2 - y_n)y + 2(z_2 - z_n)z = (x_2^2 + y_2^2 + z_2^2) - (x_n^2 + y_n^2 + z_n^2) - (d_2^2 - d_n^2) \\ \dots\dots\dots \\ 2(x_{n-1} - x_n)x + 2(y_{n-1} - y_n)y + 2(z_{n-1} - z_n)z = (x_{n-1}^2 + y_{n-1}^2 + z_{n-1}^2) - (x_n^2 + y_n^2 + z_n^2) - (d_{n-1}^2 - d_n^2) \end{cases}$$

上式可表示为：

$$AX = b$$

其中：

$$A = \begin{bmatrix} 2(x_1 - x_n) & 2(y_1 - y_n) & 2(z_1 - z_n) \\ 2(x_2 - x_n) & 2(y_2 - y_n) & 2(z_2 - z_n) \\ \dots\dots\dots \\ 2(x_{n-1} - x_n) & 2(y_{n-1} - y_n) & 2(z_{n-1} - z_n) \end{bmatrix}$$

$$b = \begin{bmatrix} (x_1^2 + y_1^2 + z_1^2) - (x_n^2 + y_n^2 + z_n^2) - (d_1^2 - d_n^2) \\ (x_2^2 + y_2^2 + z_2^2) - (x_n^2 + y_n^2 + z_n^2) - (d_2^2 - d_n^2) \\ \dots\dots\dots \\ (x_{n-1}^2 + y_{n-1}^2 + z_{n-1}^2) - (x_n^2 + y_n^2 + z_n^2) - (d_{n-1}^2 - d_n^2) \end{bmatrix}$$

则可得移动终端 X 的坐标为：

$$\hat{X} = (A^T A)^{-1} A^T b$$

通过上一步对移动终端与基站的距离修正后，得到了较为精确的距离 d ，利用极大似然法[3]，可求得终端坐标。

通过赛题所给的测试用例及对应的正确的终端位置，在利用上述模型求解过程中，对于移动终端 x 方向和 y 方向的位置，模型有很好的精确度。但是对于 z 方向，求出的移动终端的坐标存在明显的不合理的数值震荡，分析其原因发现：在实验过程中，利用极大似然法求解时需要计算线性方程组 $A^T A x = A^T b$ ，其理论解为 $x = (A^T A)^{-1} A^T b$ 。但是由于 z 坐标相对于 x, y 坐标变化不大，且 z 坐标的量级远小于 x, y 坐标的量级，其中 $A^T A$ 的条件数非常大，导致求得的 z 方向的坐标存在明显的数值震荡。

矩阵 A 的条件数定义为 A 的范数与 A^{-1} 的范数的乘积，即 $cond(A) = \|A\| \cdot \|A^{-1}\|$ ，本文中范数取 L^2 范数，条件数越大矩阵越病态。

事实上，条件数还表示了矩阵计算对于误差的敏感性。对于线性方程组 $Ax = b$ ，如果 A 的条件数大， b 的微小改变就能引起解 x 较大的改变，数值稳定性差。反之，如果 A 的条件数小， b 有微小的改变， x 的改变也很微小，数值稳定性好。条件数同样可以反映 b 不变， A 有微小改变时， x 的变化情况。

因此求解非齐次线性方程组 $A^T A x = A^T b$ 时，我们采用引入规则项的方法。对于方程组 $A^T A x = A^T b$ ，加入 L^2 规则项 $x^* = (A^T A + \beta I)^{-1} A^T b$ ，就可以直接求逆

了。

因为要得到这个解，我们通常并不直接求矩阵的逆，而是通过解线性方程的方式（例如高斯消元法）来计算。考虑没有规则项，也就是 $\beta=0$ 的情况，如果矩阵 $A^T A$ 的条件数很大时，解线性方程组就会在数值上相当不稳定，而引入这个规则项则可以改善条件数。

另外，如果使用迭代优化的算法，条件数太大仍然会使迭代算法可能不收敛，而规则项从优化的角度来看，实际上是将目标函数变成了 β 强凸的。

5.4 预估-校正-自适应定位模型

根据TOA定位原理、多基站联合定位算法、信道传输损耗模型、极大似然估计方法，本文提出了一种新的定位模型。

该模型先依据TOA定位原理预估一个近似的距离误差因子 a ，再通过信道传输损耗模型修正终端与基站之间的距离 d ，最后使用改进后的极大似然估计方法估计出精度较高的终端位置。由于本文使用的信道传输损耗模型具有自适应识别LOS和NLOS传播环境的特性，因此本文将新模型称为“预估-校正-自适应定位模型”。

该模型的算法共分三步：

第一步：预估距离误差因子 a 。即求解方程组

$$\begin{cases} (x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2 = a(vt_1)^2 \\ (x_2 - x)^2 + (y_2 - y)^2 + (z_2 - z)^2 = a(vt_2)^2 \\ (x_3 - x)^2 + (y_3 - y)^2 + (z_3 - z)^2 = a(vt_3)^2 \\ (x_4 - x)^2 + (y_4 - y)^2 + (z_4 - z)^2 = a(vt_4)^2 \end{cases}$$

其中 (x_i, y_i, z_i) 为基站的位置坐标， (x, y, z) 为某一终端的位置坐标， t_i 表示该终端与基站 i 之间的信号传播时间。由数值代数的知识和传播环境，我们知道，解是存在且不唯一的。选取距离误差因子 a 时，应注意 $0 < a \leq 1$ ，也可以多选几组基站，计算其均值作为预估值。

第二步：修正终端与基站之间的距离 d 。详见信道传输损耗模型的介绍。

第三步：使用改进后的极大似然估计方法估计出精度较高的终端位置。

该算法的流程图如图8所示。

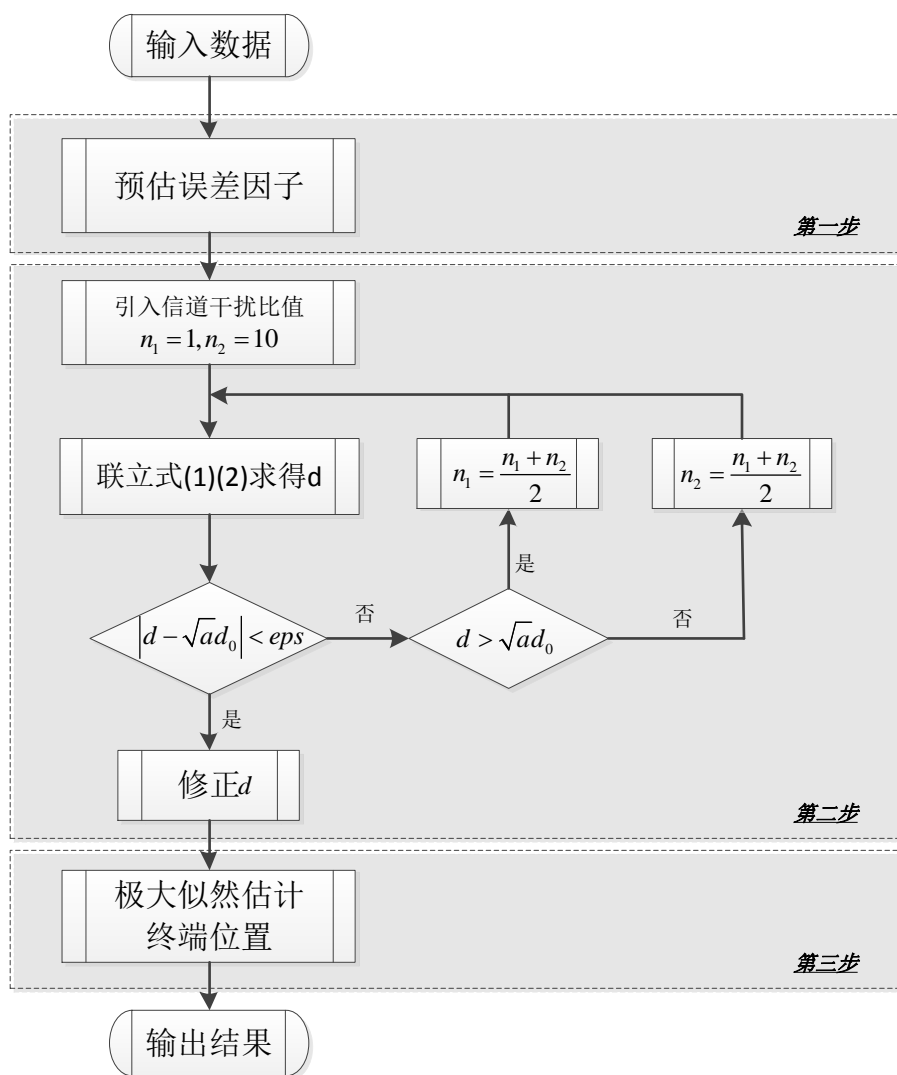


图 8 算法流程图

六、问题分析与求解

6.1 问题一的分析与求解

本题给出 10 组 LOS 或 NLOS 传播环境下从手持终端到基站的 TOA 测量数据和所有基站的三维坐标，要求根据这些测量数据计算出终端的三维坐标。

首先，本文利用预估-校正-自适应定位模型，直接利用未改进的极大似然估计求得的结果如图 9 和图 10 所示。

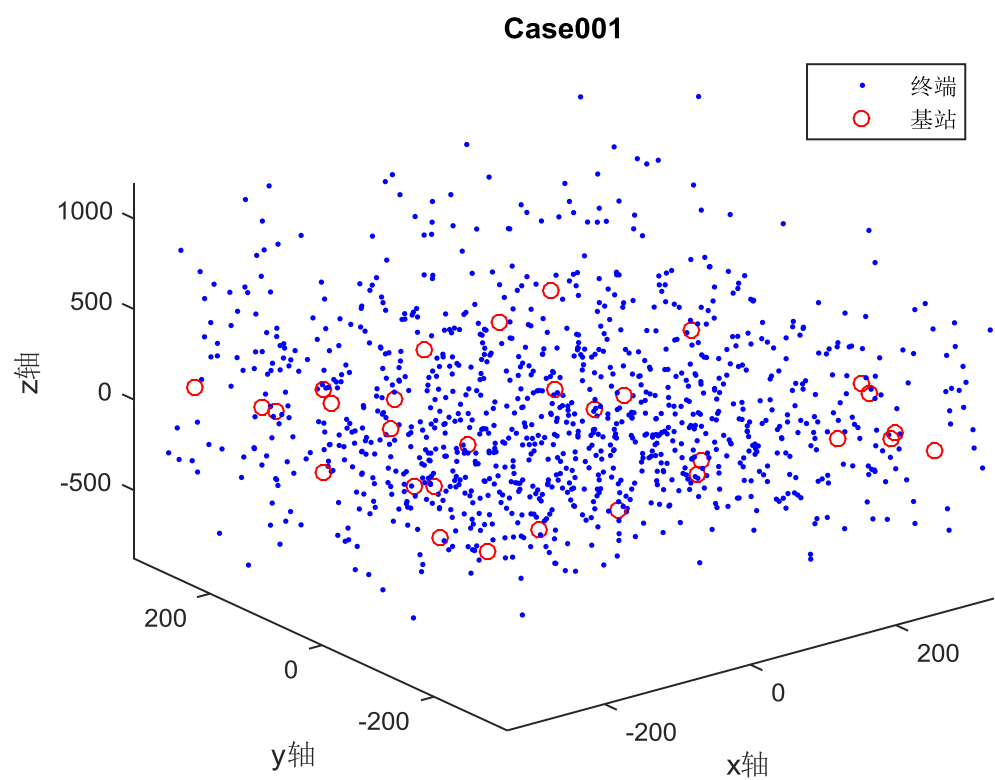


图 9 Case001 的终端三维坐标

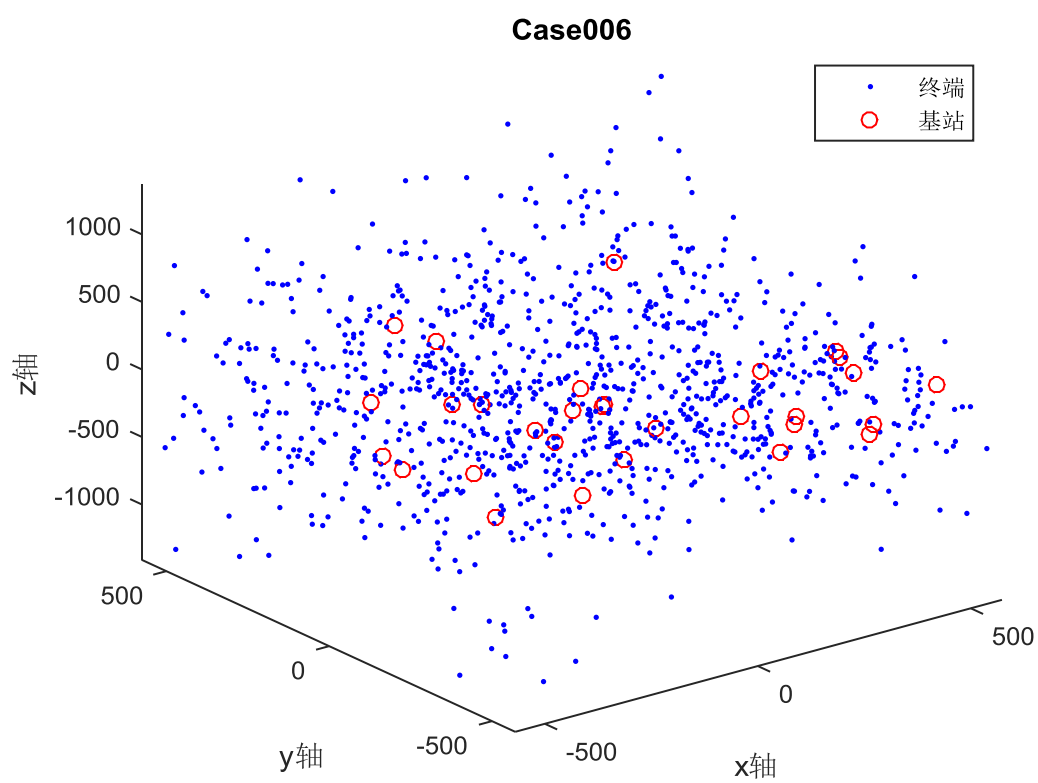


图 10 Case006 的终端三维坐标

其他结果详见附录一和附录二。
利用改进后的预估-校正-自适应定位模型计算的结果,如图 11 和图 12 所示。

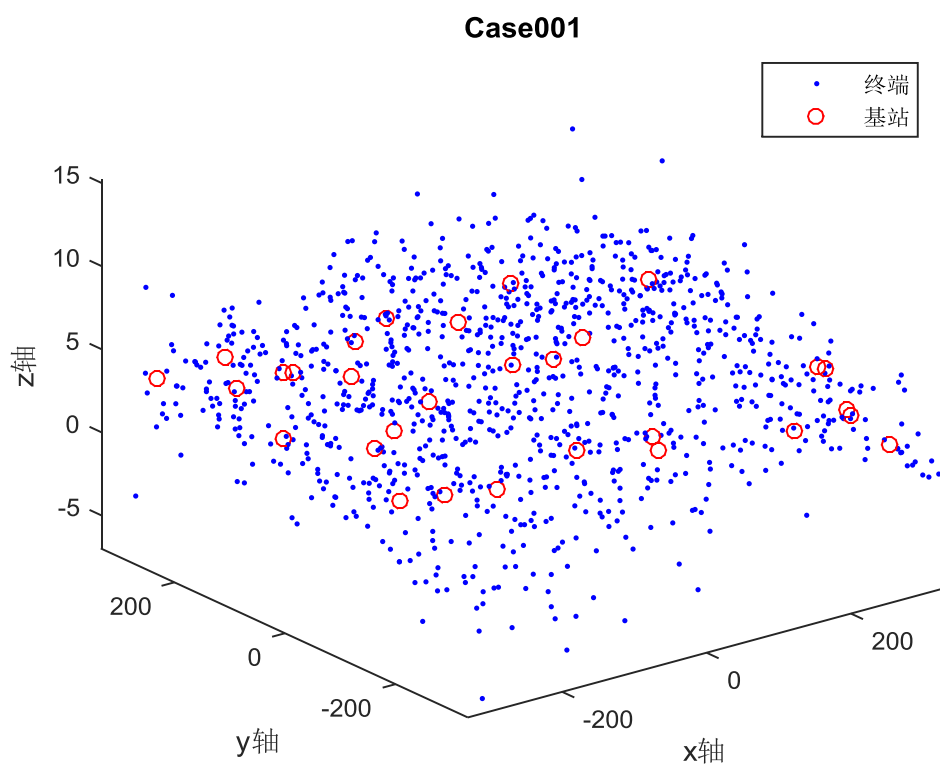


图 11 Case001 的终端三维坐标

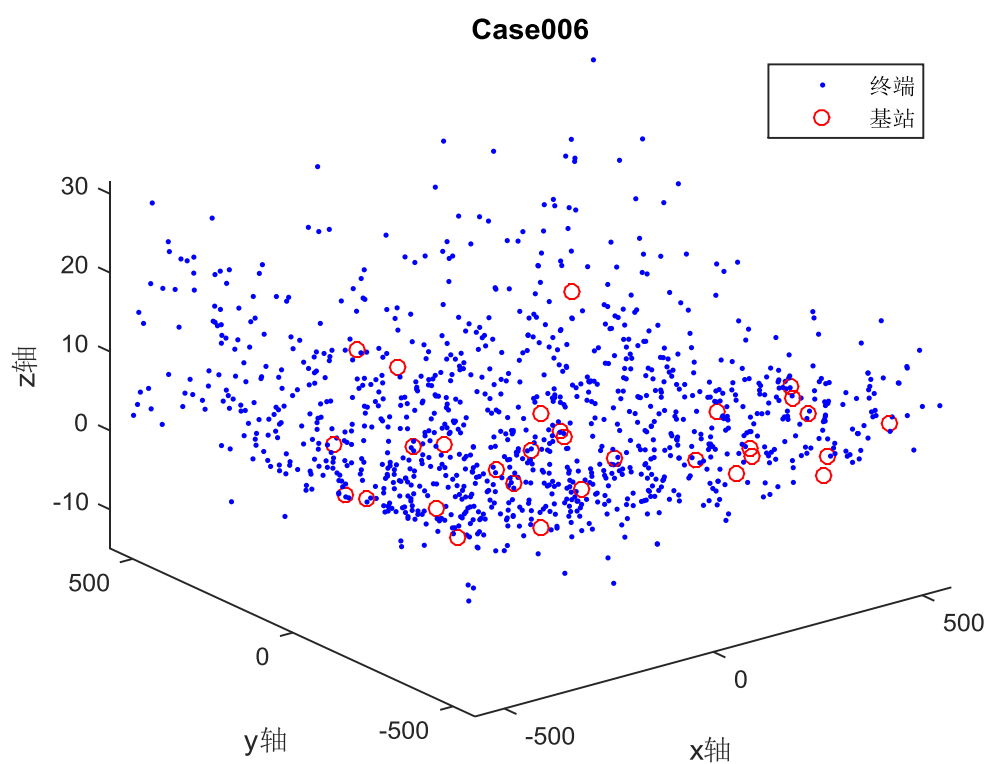


图 12 Case006 的终端三维坐标

在对模型进行优化后，我们将优化后的算法再次用到测试用例中，将计算的移动终端的坐标与正确的终端位置比较。*RMSE* 的结果如下：

表 2 模型优化前后 *RMSE* 对比

测试场景	优化前	优化后	比率
001	294.1350	6.8822	42.7
002	571.6403	12.3531	46.3
003	220.4731	5.0992	43.2
004	190.2598	4.8522	39.2
005	446.8602	4.2634	104.8

表 2 表明优化后的 *RMSE* 显著减小，即优化后的模型更适合应用到三维定位领域。

6.2 问题二的分析与求解

问题二要求使用尽可能少的基站数目，实现近似最优的三维定位精度。由于定位一个终端至少需要 4 个 *TOA* 值，对于给出的 10 个 case 场景，我们从每一个 case 中的全部基站数中分别随机选取 $i (i = 4, \dots, 10)$ 个基站数，分别进行了 20 组定位实验并得到手持终端坐标的 *RMSE* 值：

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{o,i} - X_m)^2}{n}}$$

其中， X_m 是建立在第一问的模型上，利用 30 个基站求得的同一终端的坐标，在第一问中已经证实利用全部基站信息计算出的终端位置比较精确，这里将其作为精确解。 $X_{o,i}$ 是利用 i 个基站分别进行 20 组实验后，每个实验得到的所有终端坐标。对比这几组实验，在保证较好精度的前提下，我们选出最优基站的数目。

本文利用构建的预估校正模型，对于同一场景，选取不同基站数进行定位。对得出的终端坐标 *RMSE* 值进行比较，结果如图 13 所示。

通过图 13 我们发现，当基站数 $n = 4$ 时，模拟出的终端坐标的 *RMSE* 值很大，即坐标波动较大，精度较低。当基站数逐渐增大时，*RMSE* 值逐渐较小，且当 $n = 5, 6, 7$ 时，减小趋势较为明显。当 $n = 7$ 时，模拟出的终端三维精度误差控制在 5 m。继续增加基站数，当 $n = 8, 9, 10$ 时，较之 $n = 7$ 时模拟出的终端坐标 *RMSE* 值相差无几，即精度无明显提升。因此我们选择 7 或 8 个终端数，来实现近似最优的三维定位精度。具体的 *RMSE* 值如表 3 所示。

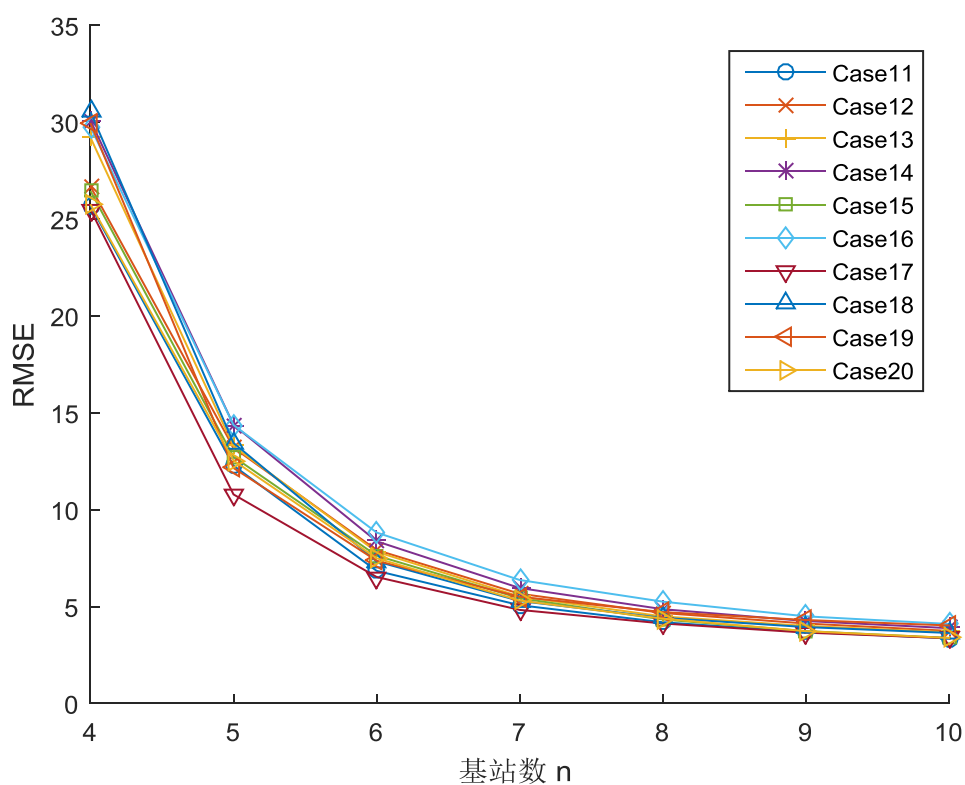


图 13 不同场景不同基站数的 $RMSE$ 值

表 3 不同场景不同基站数的 $RMSE$ 值

基站数 (n) 场景	4	5	6	7	8	9	10
011	25.6959	12.2770	6.8462	5.0724	4.1918	3.6903	3.3628
012	26.6828	13.1985	7.9213	5.6638	4.6702	4.1272	3.7475
013	29.2123	13.2793	7.8266	5.5207	4.4929	4.0017	3.6751
014	30.0880	14.3778	8.3723	5.9597	4.8668	4.2560	3.8882
015	26.4568	12.7330	7.6323	5.4347	4.3350	3.7372	3.3830
016	29.7563	14.3715	8.8239	6.3732	5.2551	4.4981	4.1045
017	25.4929	10.7902	6.5295	4.8298	4.1293	3.6569	3.3678
018	30.5338	13.4450	7.3358	5.2889	4.4198	3.9435	3.6402
019	29.9564	12.1537	7.3676	5.5066	4.7176	4.3076	4.0267
020	25.7660	12.5415	7.4969	5.3123	4.3790	3.7307	3.3813

6.3 问题三的分析与求解

本题给定了 5 组处于移动过程中的终端采集到的 TOA 数据，与问题二的求解过程类似，利用本文提出的预估-校正-自适应定位模型求解。为了降低计算复

杂度加快计算速度，我们利用第二问的结果，选取 7 个基站作为参考基站来实现对移动过程中的终端进行二维定位。得出结果后，我们拟合出终端的移动轨迹，并计算出相应的轨迹方程。

计算出移动中的终端的二维坐标结果如下散点图所示。我们发现各个终端的运动轨迹大致上是二次曲线，并用二次拟合得到了这些终端的运动轨迹，拟合出的轨迹如图 14 所示。

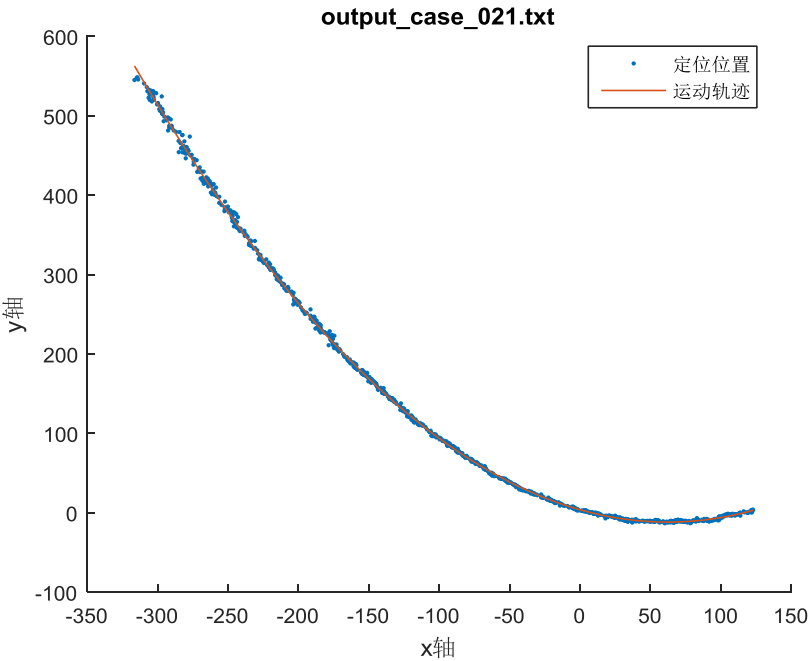


图 14 Case021 运动轨迹

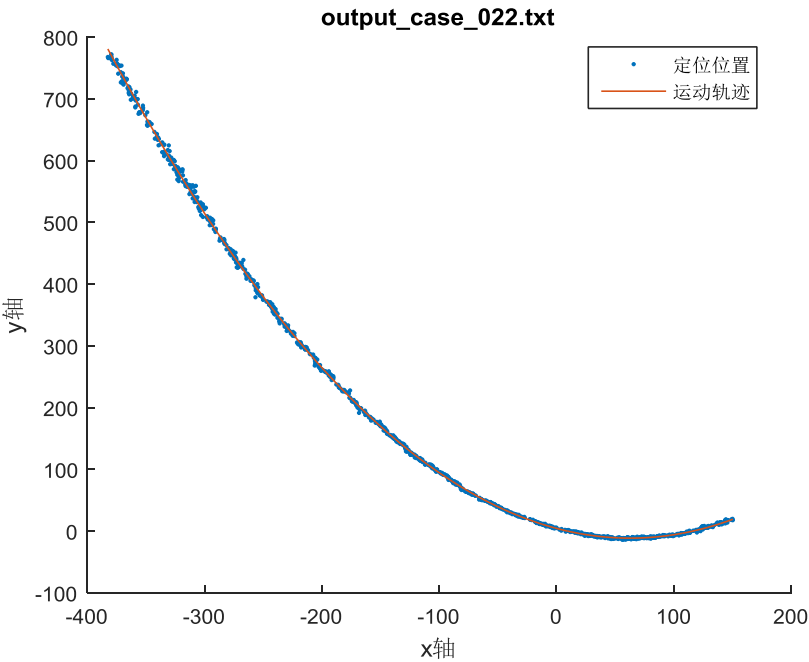


图 15 Case022 运动轨迹

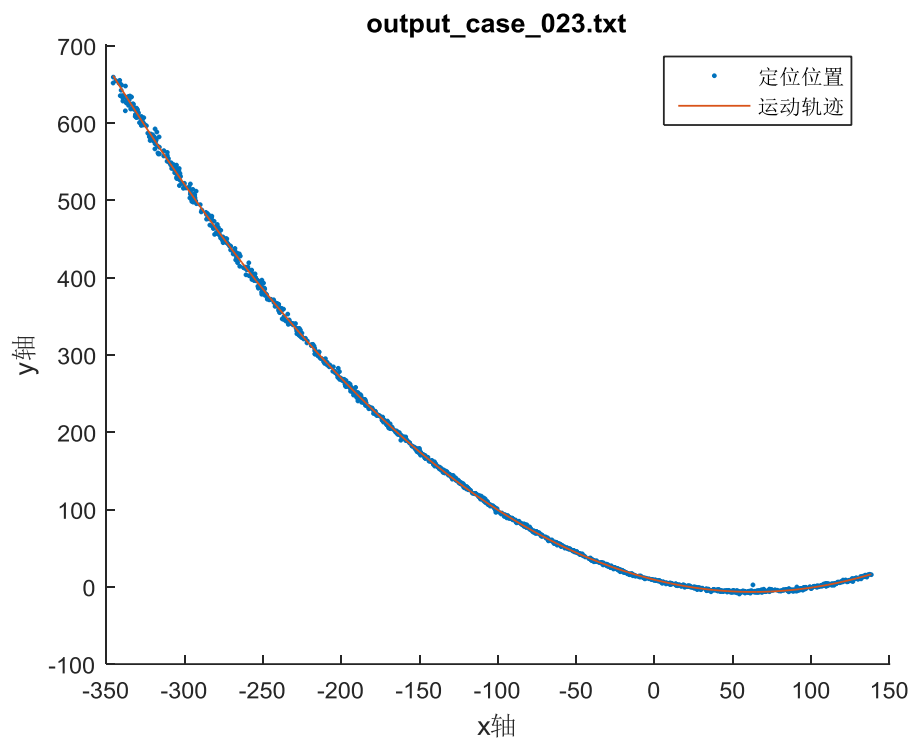


图 16 Case023 运动轨迹

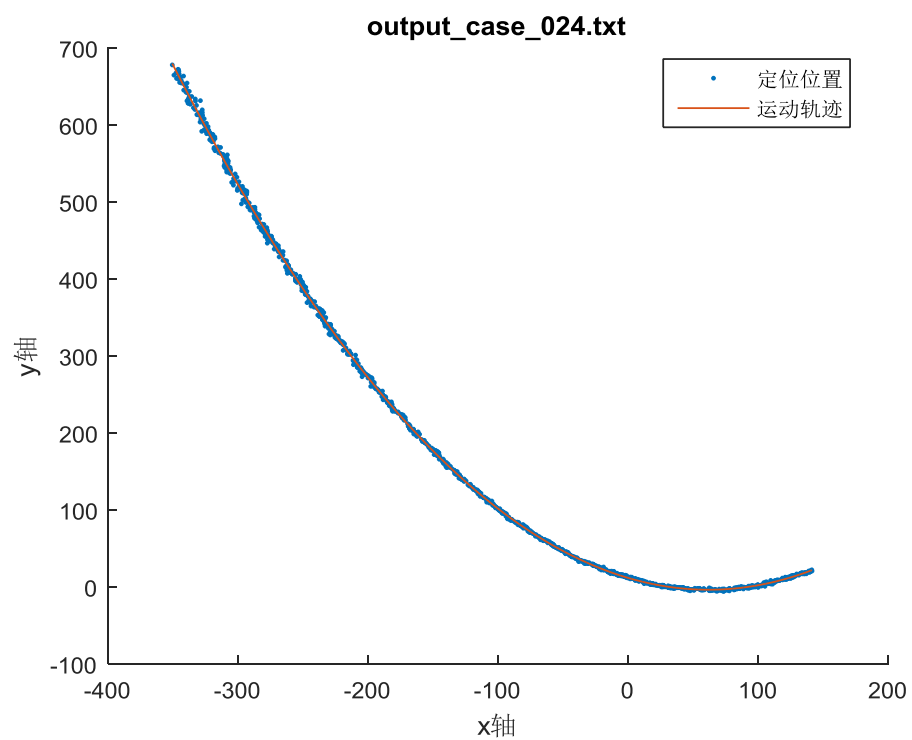


图 17 Case024 运动轨迹

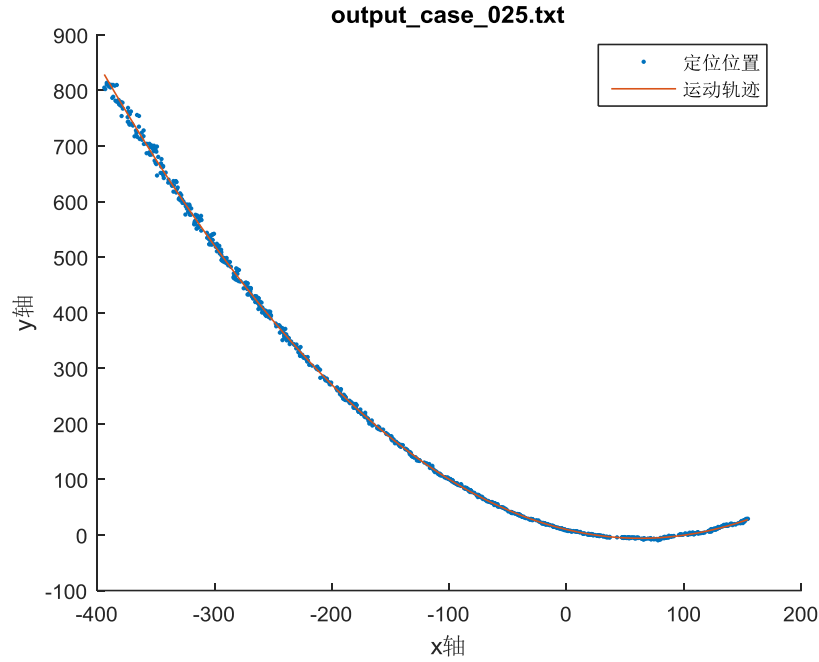


图 18 Case025 运动轨迹

表 4 各个终端的轨迹方程

终端名	轨迹方程	范围	拟合误差
Case021	$y = 0.0040x^2 - 0.49923x + 3.9532$	$[-316.5307, 123.0845]$	24.0216
Case022	$y = 0.0040x^2 - 0.5025x + 4.6274$	$[-382.5530, 150.4715]$	22.8729
Case023	$y = 0.0040x^2 - 0.5023x + 9.1369$	$[-345.7545, 138.5132]$	20.2206
Case024	$y = 0.0040x^2 - 0.4983x + 12.2384$	$[-350.5834, 141.6730]$	21.6814
Case025	$y = 0.0040x^2 - 0.5028x + 9.8873$	$[-394.1005, 154.8275]$	30.0027

从上述图表看，本文对各个终端的轨迹方程的拟合误差在 20 ~ 30 之间，拟合结果较好。

6.4 问题四的分析与求解

第四问中，考虑到实际物理世界中基站的通信半径有限，且假设通信半径为 200 米。本题通过测得的移动终端与基站的 TOA 数据，利用模型中第一步的初估，计算出终端与基站的距离，进而筛选出小于 200 米的数据。根据选出的有效数据，我们统计各个终端可用的 TOA 数据的数目（即连接数），由于每个终端实现定位需要被至少 4 个基站检测到，所以我们对连接数大于等于 4 的终端进行定位。进行定位计算时，我们仍采用前述的预估-校正-自适应模型来对终端进行精确定位。

文中要求建立模型分析连接度数 λ 与定位精度之间的关系。我们采用如下的定位精度定义：

$$LAC = \sqrt{\frac{n}{\sum_{i=1}^n (X_{o,i} - X_m)^2}}$$

其中， n 为可定位终端的近似坐标的个数（例如某终端只能被 5 个基站检测到，则可求的近似坐标个数为 $C_5^4 = 5$ ）， $X_{o,i}$ 为模型所求的近似坐标， X_m 是用所有可连接的基站算出的较为精确的终端坐标。 LAC 指标的意义即为定位的误差越小，精度数越大。求得可以被定位的终端的位置与连接数后，利用题中所给的平均连接度数的定义求得 λ ，最终绘制出连接度数 λ 与定位精度的关系曲线。

计算出不同场景下可以被定位的终端的标号及坐标见附件。不同场景下求得的可以被定位终端到基站的连接数与精度的关系如下图所示。

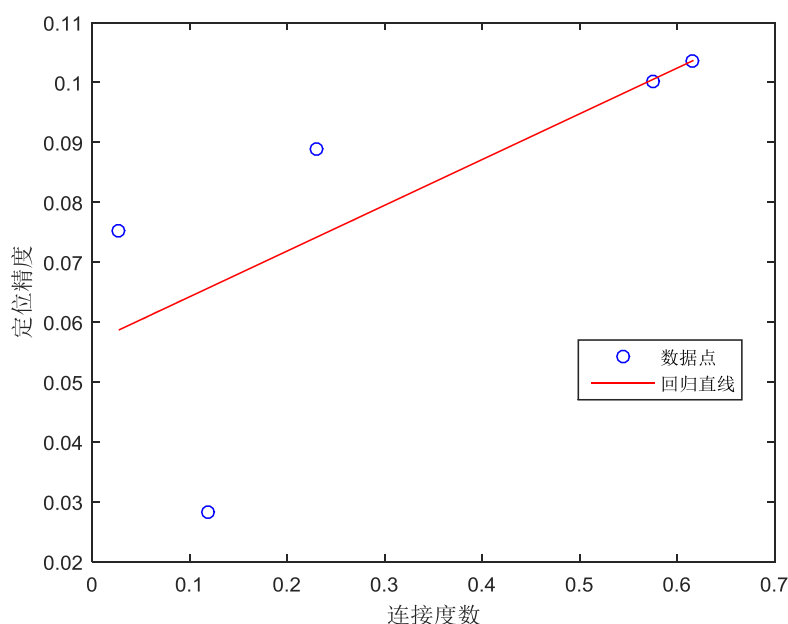


图 19 连接度数与定位精度的关系

回归方程为： $LAC = 0.0764\lambda + 0.0566$ 。

从图中可以看出，定位精度随着连接度数 λ 的增加而增加，即当一个移动终端能够被更多的基站定位时，运用我们的模型计算出的终端坐标将更精确，这个结论也符合第二问的结果。具体的连接数、连接度数与精度值见下表。

表 5 各场景连接度数与精度

场景	连接数	终端数	连接度数	精度
Case026	130	1100	0.1182	0.0283
Case027	739	1200	0.6158	0.1035
Case028	299	1300	0.2300	0.0890
Case029	804	1400	0.5743	0.1002
Case030	27	1000	0.027	0.0752

七、模型的不足与展望

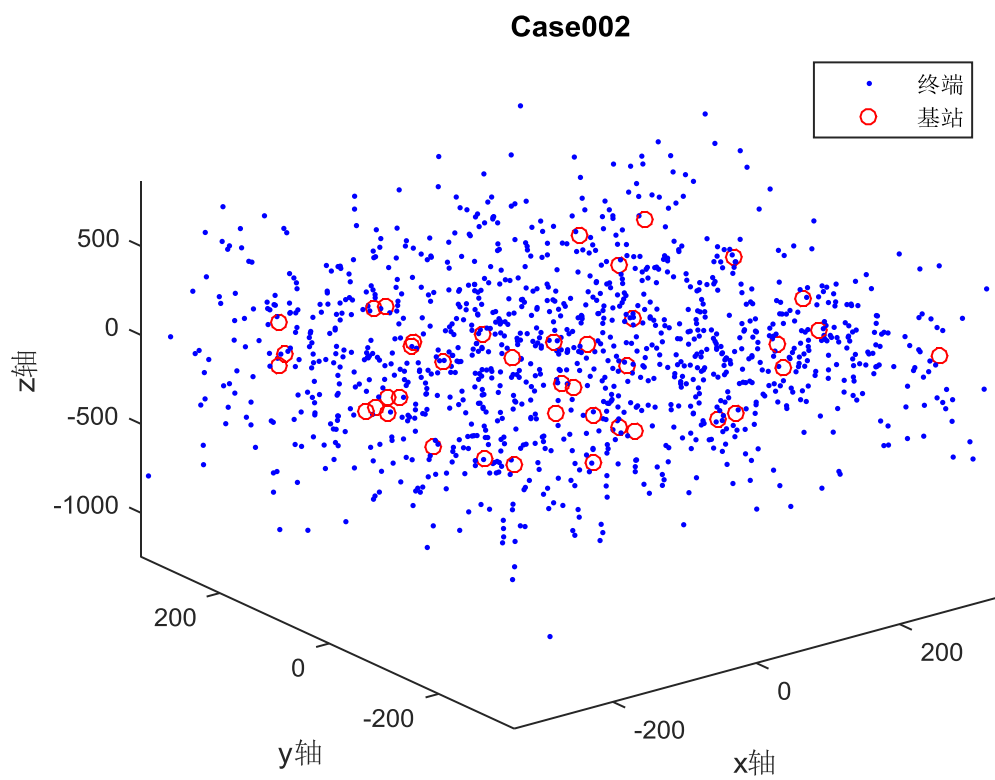
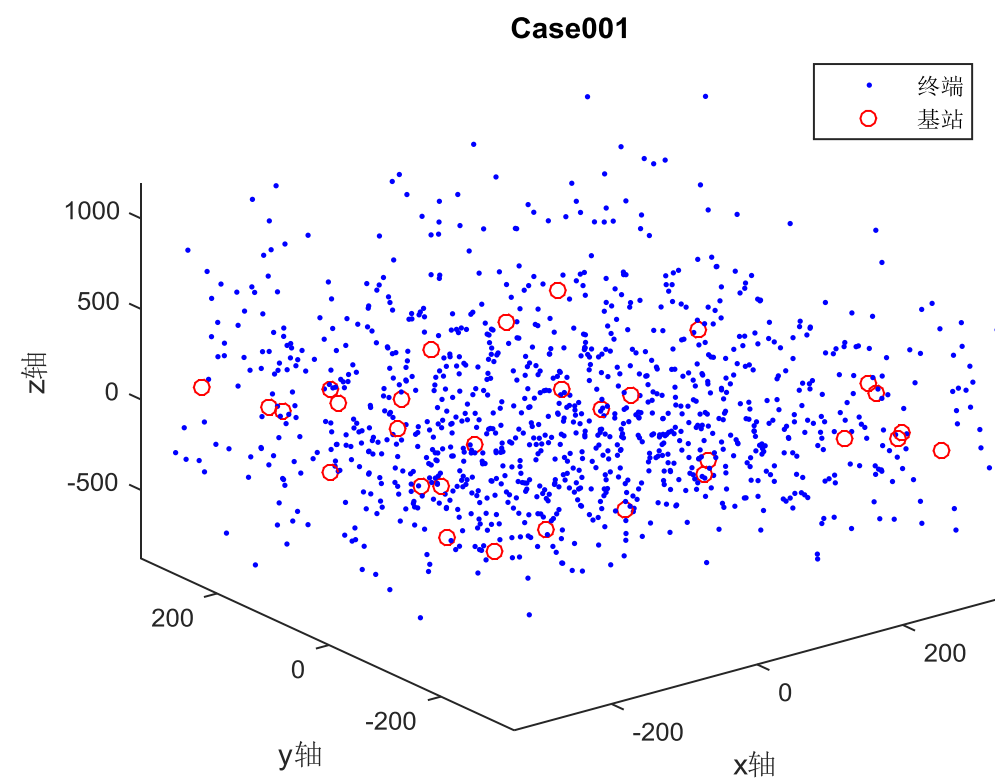
1. 本文初探了基于无线电信号传播的预估-校正-自适应定位模型，但所考虑的环境比较简单，未能细致地考虑障碍物对信号的反射、折射而引起的信号变化，且室内物品的摆放位置往往不定，室内布局也会时常变化，使用固定的距离误差因子的模型就无法适用于这些环境，因此可从距离误差因子出发进一步改善本文的定位模型。
2. 模型中估计距离误差因子时使用了解析的方法，求解速度慢，今后需简化计算，提高算法的效率。
3. 模型中应对病态的正规方程组的求解需进一步研究改善。
4. 模型是建立在 TOA 方法的基础上的，且考虑了 NLOS 的影响。而在实际应用中，除 TOA 定位方法外，还有如 TDOA 等其他定位方法值得研究。

参考文献

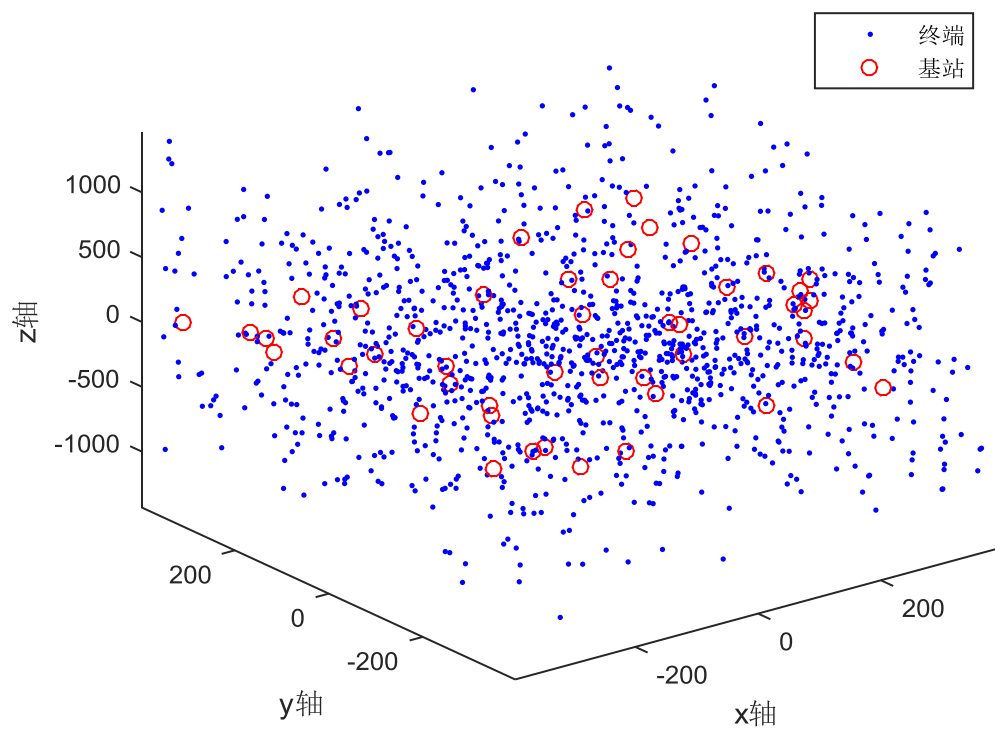
- [1] 魏晓莉. 减小 NLOS 影响的 TOA 定位算法及其在 GSM 网中的应用研究[D]. 湖南大学, 2005.
- [2] 李辉. 基于 RSSI 的高精度室内三维空间定位算法[D]. 河南大学, 2011.
- [3] 刘德亮. 室内环境下无线定位关键技术研究[D]. 天津大学, 2015.
- [4] 贺黎滔. 基于移动通信网络的高精度定位关键技术研究[D]. 北京邮电大学, 2010.
- [5] 赵永翔. 基于无线局域网的室内定位系统研究[D]. 武汉大学, 2010.
- [6] 张宴龙. 室内定位关键技术研究[D]. 中国科学技术大学, 2014.
- [7] 陈豫章. 基于 RSSI 的室内三维定位技术研究与实现[D]. 云南大学, 2013.
- [8] 李婧. RFID 室内三维定位算法研究[D]. 电子科技大学, 2013.
- [9] 王昕, 王宗欣, 刘石. 一种考虑非视线传播影响的 TOA 定位算法[J]. 通信学报, 2001, 03:1-8.
- [10] 刘林. 非视距环境下的无线定位算法及其性能分析[D]. 西南交通大学, 2007.
- [11] 张明华, 张申生, 曹健. 无线局域网中基于信号强度的室内定位[J]. 计算机科学, 2007, 06:68-71+75.
- [12] 杜超, 韦再雪, 常永宇, 桑林. 室内定位新技术浅析[J]. 信息通信技术, 2015, 05:71-77.

附 录

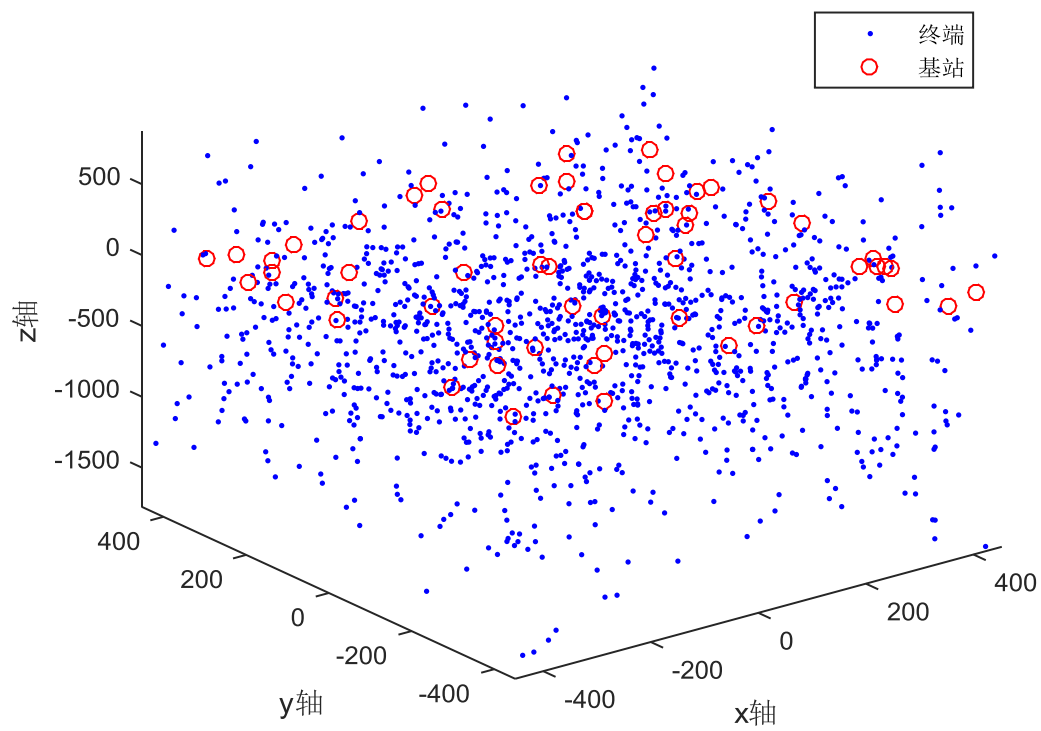
附录一 问题一改善前的计算结果图



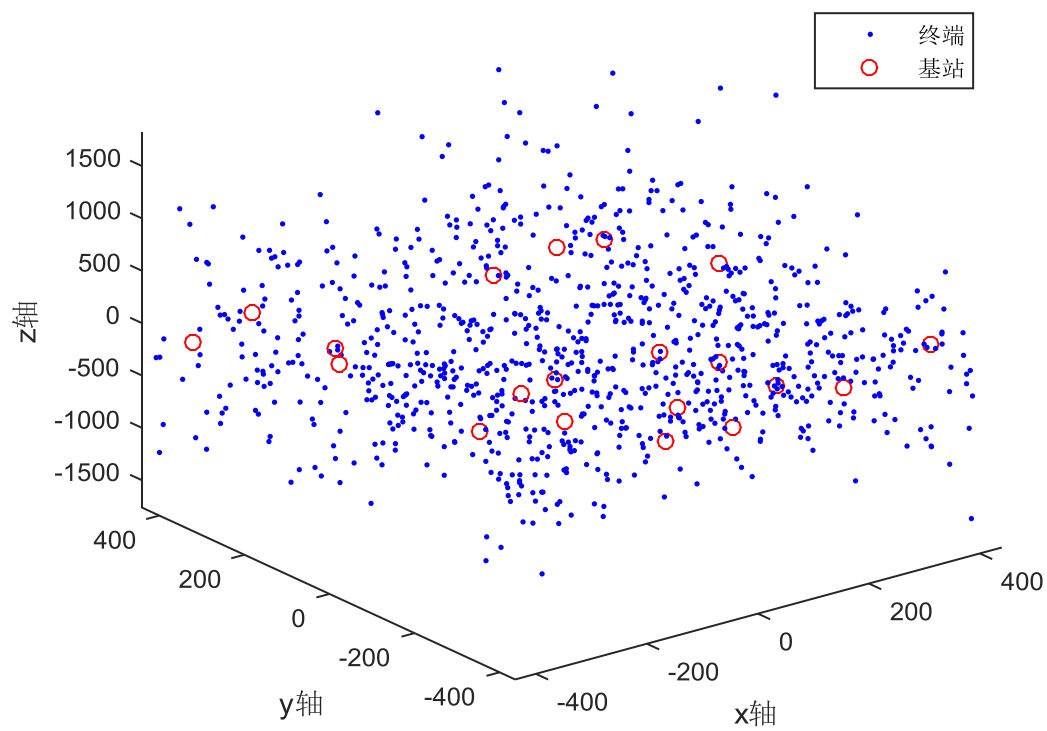
Case003



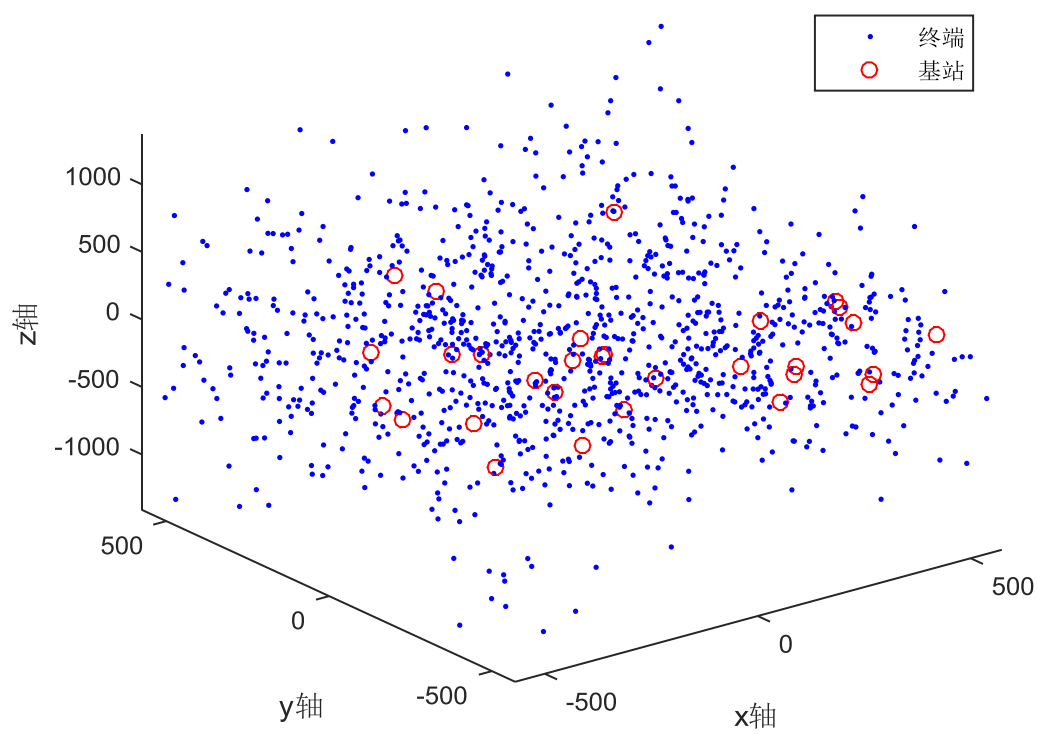
Case004



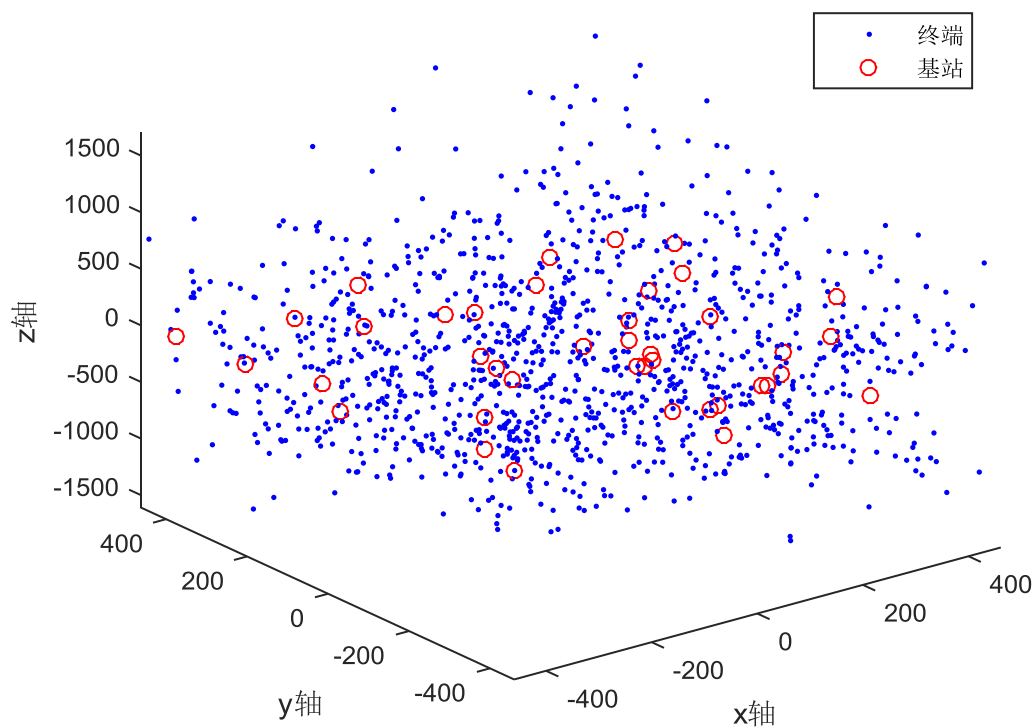
Case005



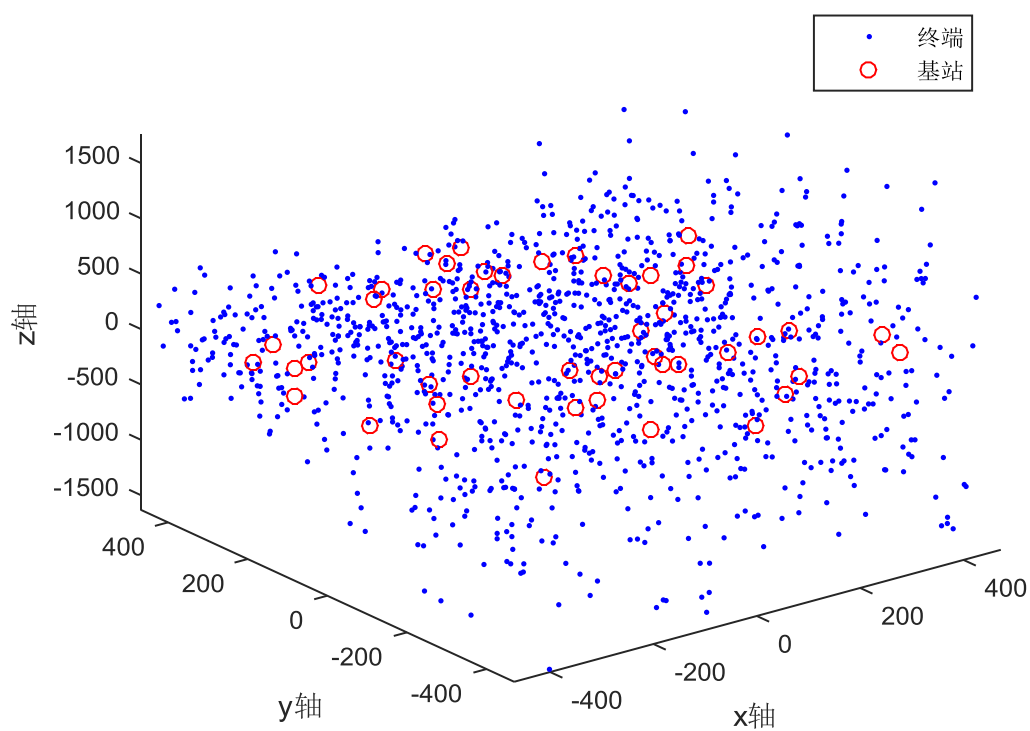
Case006

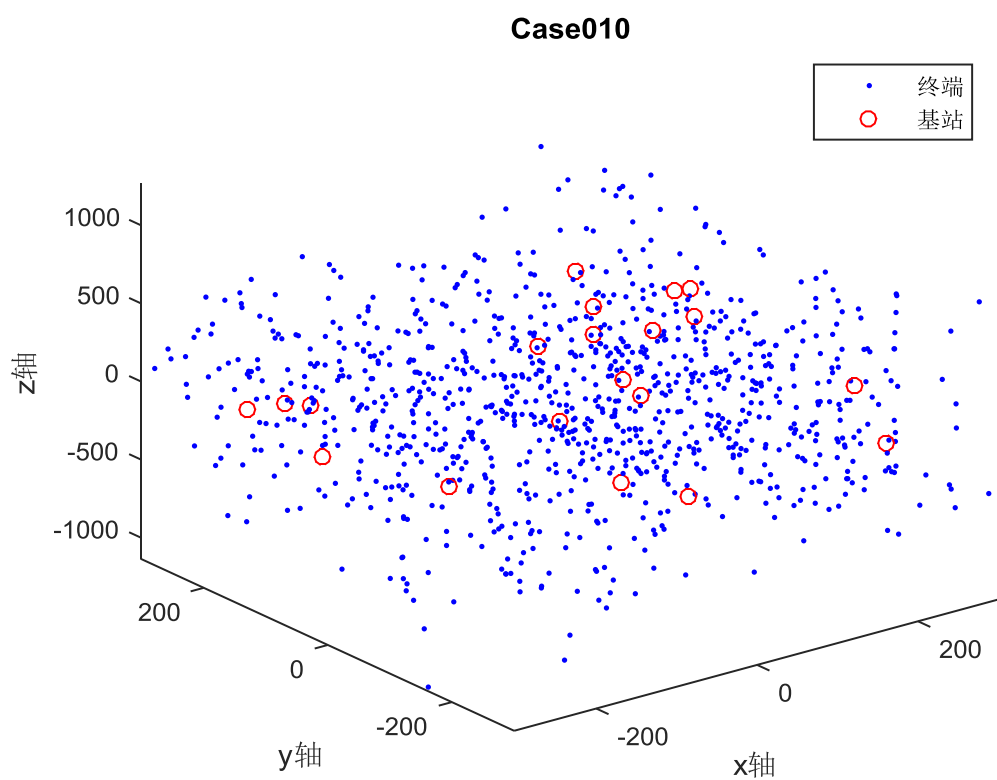
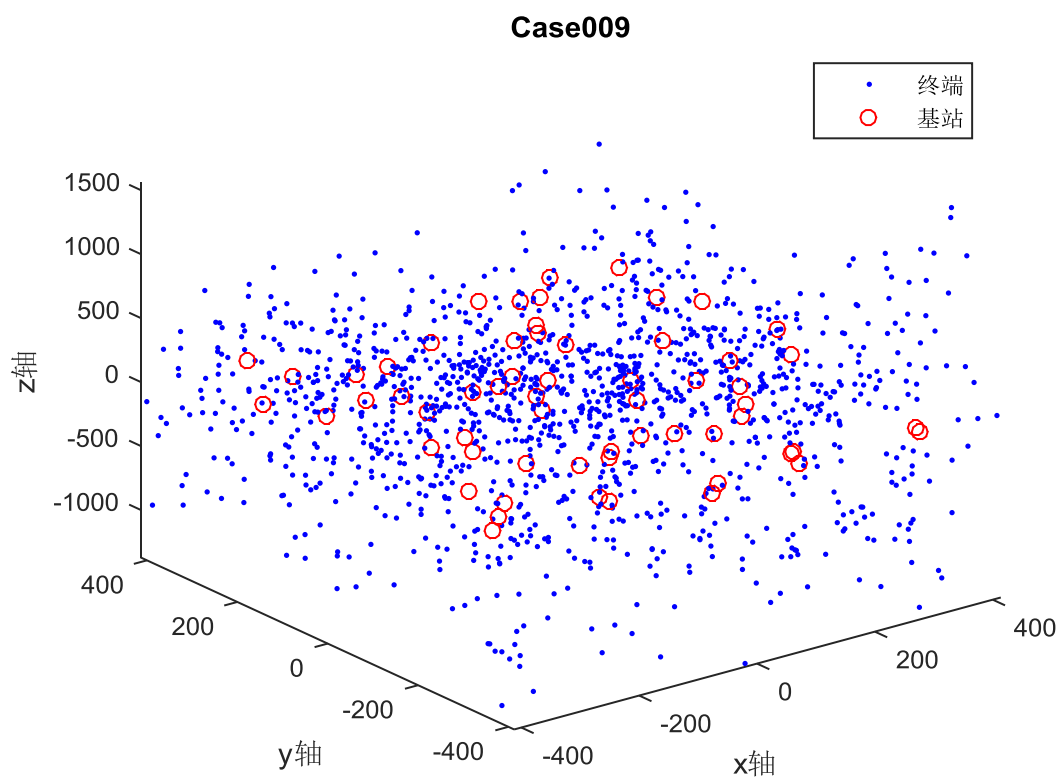


Case007



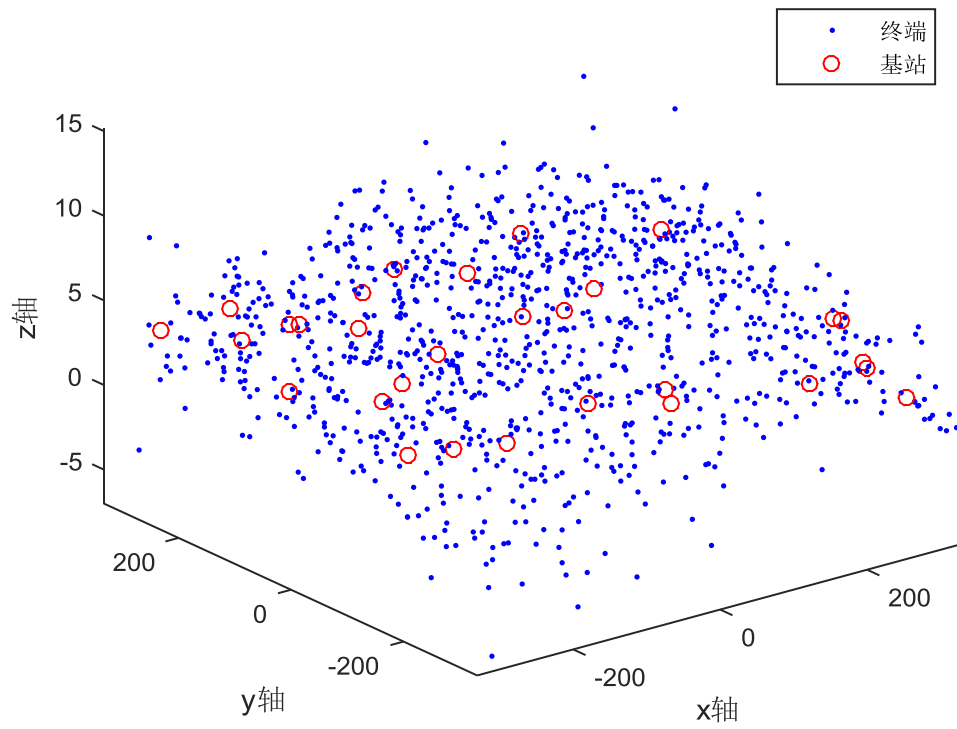
Case008



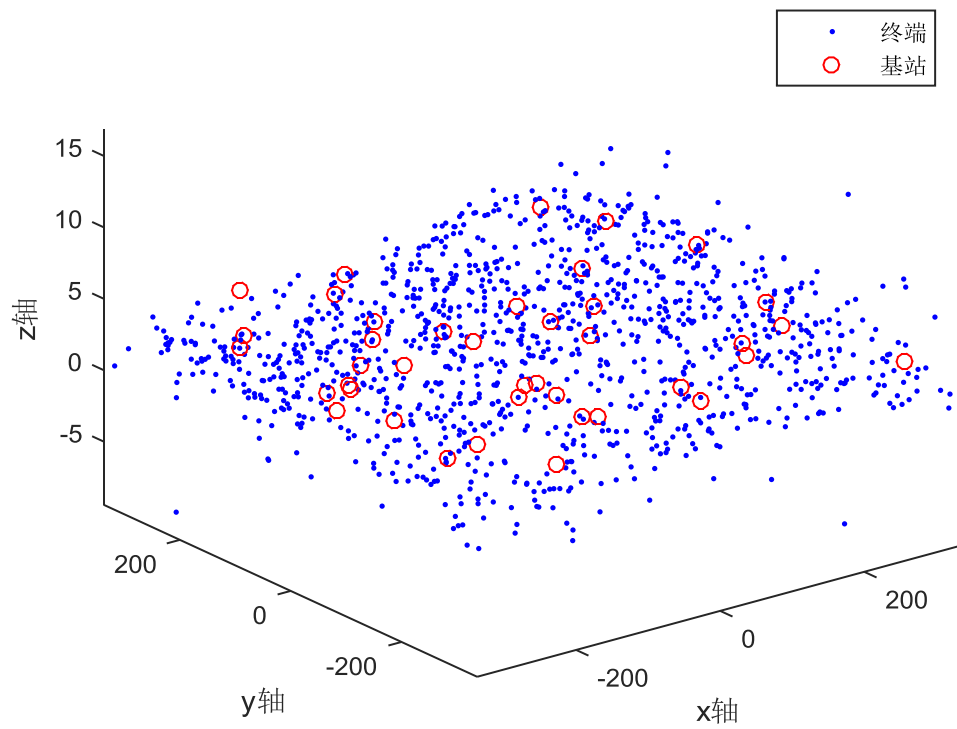


附录二 问题一改善后的计算结果图

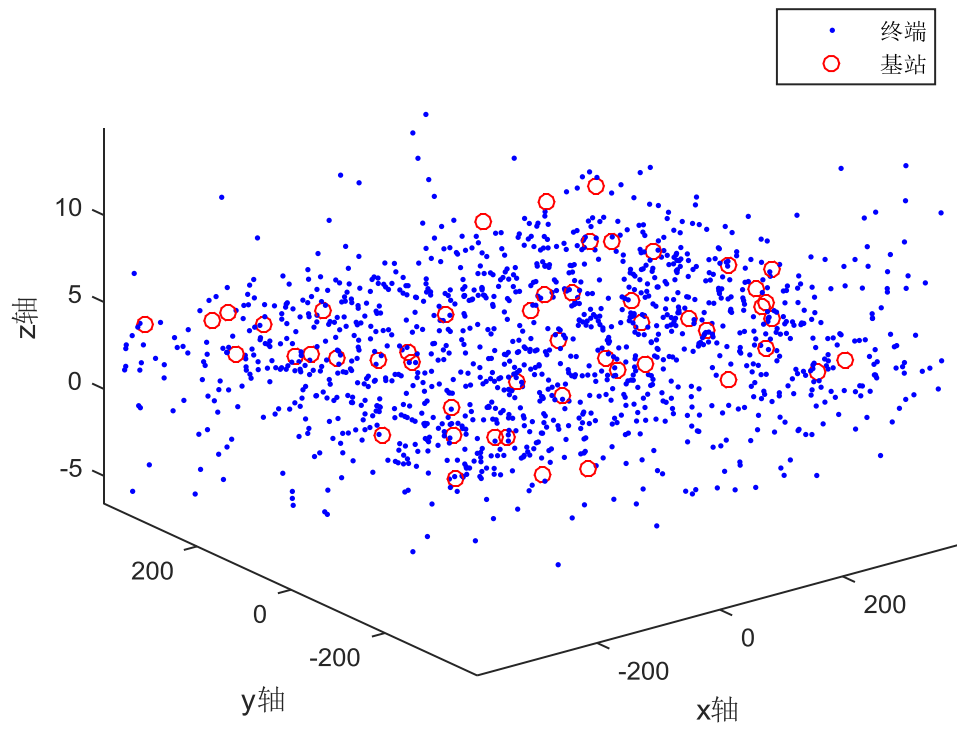
Case001



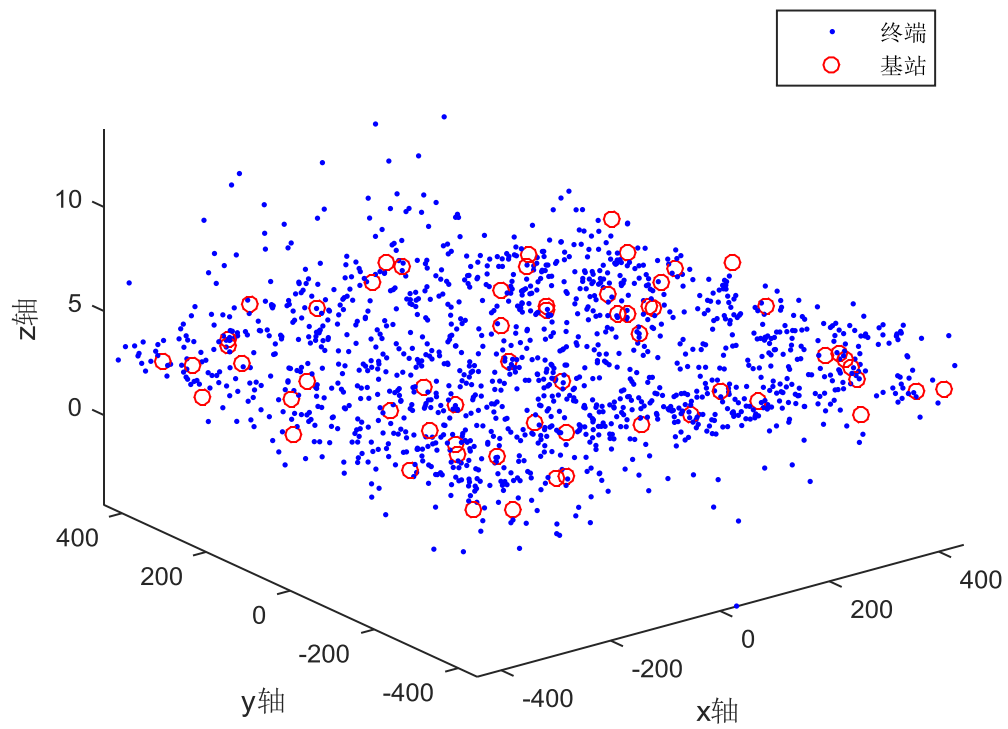
Case002

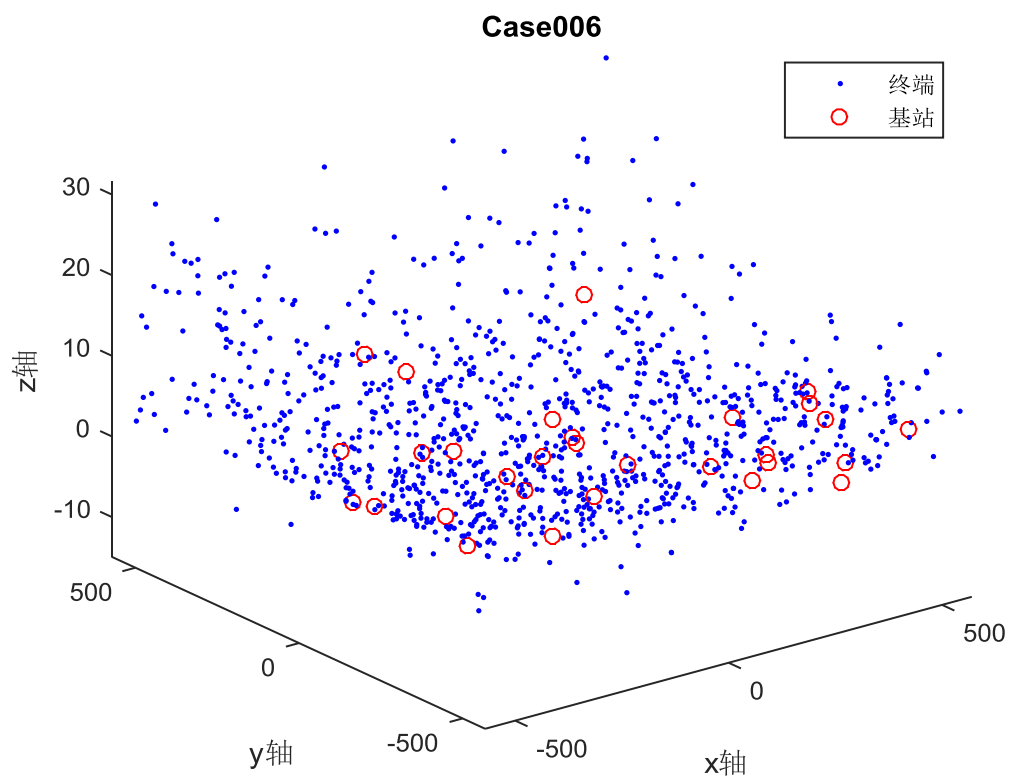
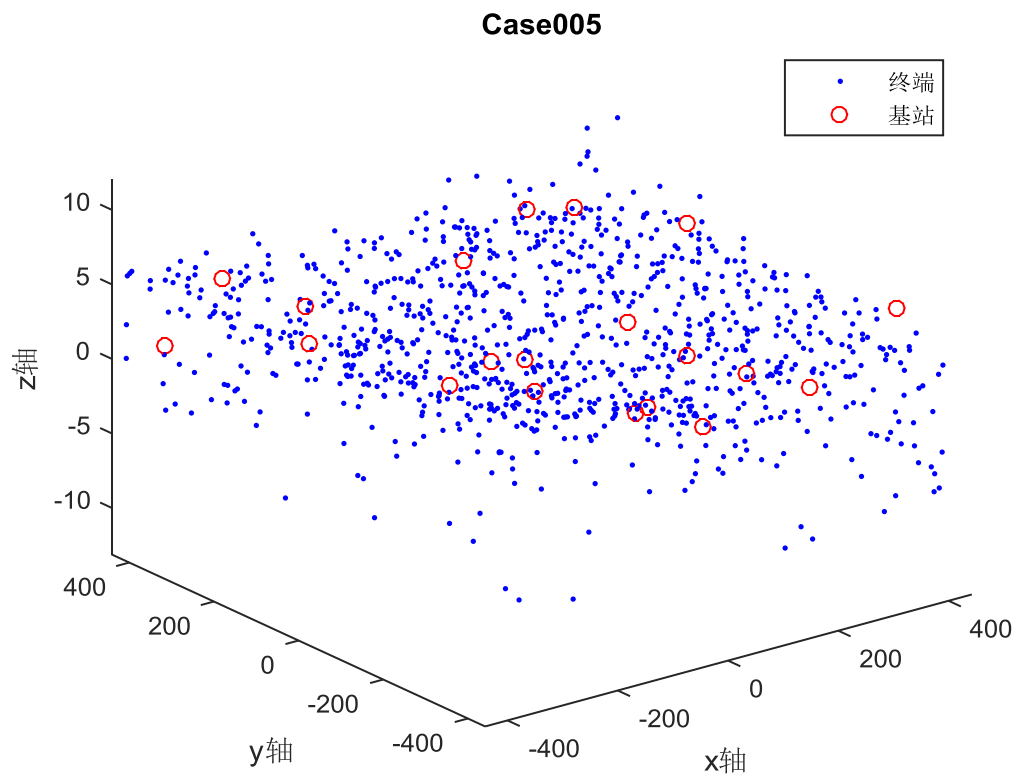


Case003

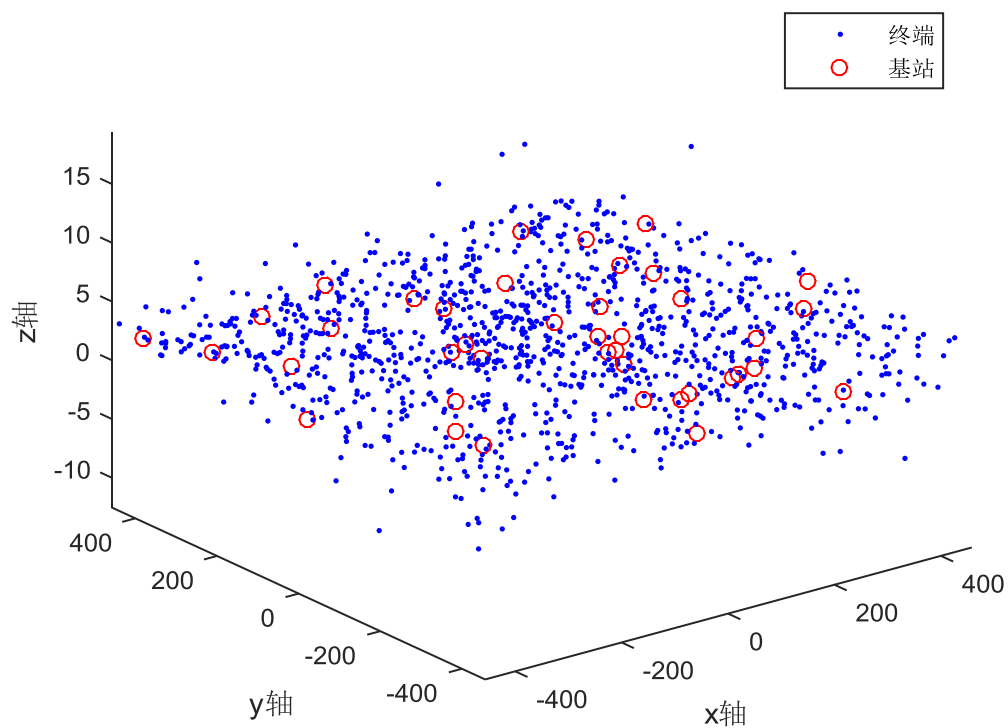


Case004

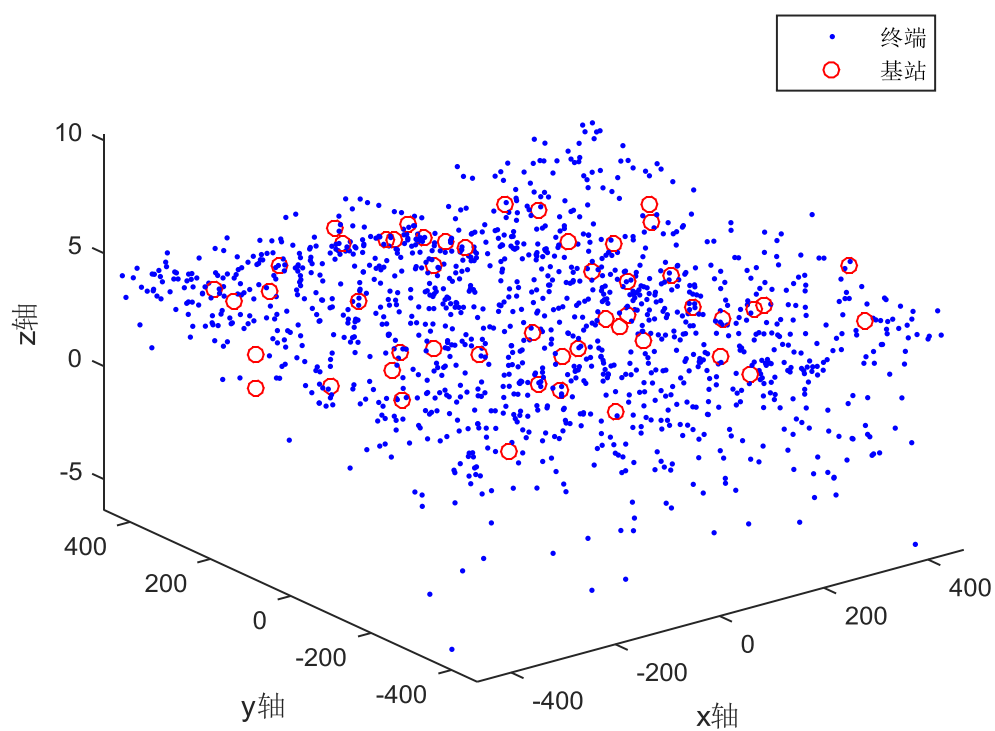




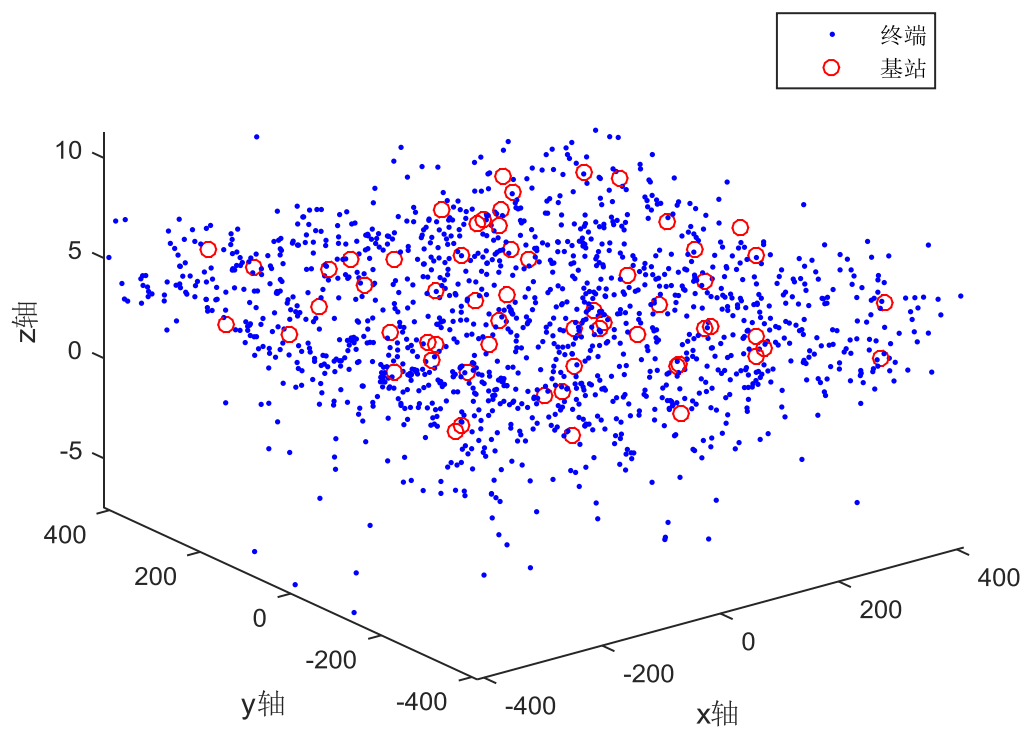
Case007



Case008



Case009



Case010

