

# 城镇天然气家庭用量分析与室内燃气泄漏安全状况模拟

## 摘要

在本文中,我们基于给出的数据分析了我国近年来燃气产业的发展情况,讨论了天然气家庭用量与数据中的各类因素的关系,并对于我国未来的燃气产业的发展趋势做出了合理预测。之后,我们基于数据确定了天然气的扩散模型,通过推导模型的参数 $t$ 得到了泄漏过程中室内的不同位置在不同时间的天然气浓度。

针对问题一,由于数据的维度较多且含有不确定因素,我们使用了灰色系统预测的方法来预测城镇的燃气普及率,同时使用主成分分析的方法进行降维,分析人工煤气的家庭使用量的变化趋势。得到大约在2035年城镇的燃气普及率达到100%,而在2019年结束人工煤气的使用历史。

针对问题二,在相关性分析部分,我们首先使用Pearson相关系数进行分析,之后选定了相关系数的绝对值 $r$ 大于0.75的6个变量,对这些变量排除天然气管道长度的影响进行偏相关分析,最终选取了偏相关系数的绝对值大于0.4的煤气家庭用量( $r_p = 0.41$ )和天然气管道长度(排除煤气家庭用量的影响, $r_p = 0.95$ )两个变量。考虑到两个变量存在一定的相关性,我们使用岭回归确定了两个变量与天然气家庭用量的关系。之后,我们使用最小二乘回归确定了以上两个变量与时间的关系,并综合两个模型确定了天然气的家庭用量与时间的关系。并基于这一模型进行了对于2025、2030、2035年的天然气的家庭用量的预测,预测结果是4523376、6373522、8546438万立方米。针对问题三,我们调整了问题二中关于天然气管道长度与时间关系的函数,使用与之前预测相同的方法得到了在这种情况下在2025年的天然气用量为6165018万立方米。

针对问题四,考虑到天然气的扩散过程的不同阶段以及天然气在扩散时延着不同方向的浓度分布,我们依托数据,通过非线性回归方法建立了模拟泄漏过程中的扩散分布的数学模型,并使用最小二乘法反向推导了模型中的参数与时间的关系。从而确定了天然气泄漏过程中天然气的时空分布。在此基础上,可以求解得到在无风的情况下在52.64s后的天然气浓度会达到触发报警器的阈值 $3.5 \times 10^{-3} \text{kg/m}^3$ 。

**关键字:** 灰色系统预测 主成分分析 相关性分析 非线性回归

## 一、问题背景与问题重述

### 1.1 问题背景

城镇燃气主要包括人工煤气、天然气和液化石油气三类，其中天然气有着清洁、热值高的优秀特性，根据附件的统计数据，我国近年来的天然气供气总量、家庭用量以及用气人口数增长迅速，而人工煤气的使用量则在逐年下降，人工煤气将会很快被优质清洁的天然气所取代。

城市燃气管道因老化、腐蚀、管材和焊缝缺陷等因素，可能导致泄漏事故，进而引发火灾及爆炸，造成人员伤亡和环境污染。因此，燃气管道安全问题受到高度重视。另外，由于我国的城市化率进一步提高，针对城市楼房室内燃气泄漏的安全问题研究也显得尤为重要。

### 1.2 问题重述

首先我们需要参考问题背景和附录中的数据，建立合适的数学模型来预测我国未来的燃气化发展趋势，并解决以下问题：

1. 城镇的燃气普及率最早会在哪一年实现 100%?
2. 排除人为干预或政策引导等因素，请问家庭使用人工煤气的历史最有可能在哪一年结束?

管道铺设长度的逐年增加和人工煤气的家庭用量的逐年减少等因素都可能导致天然气供气量逐年加大，因此我们需要充分考虑各类因素，分析数据中各个因素间的相关性，并构建天然气家庭使用量连续增长模型，同时利用这一模型对 2025 年、2030 年和 2035 年全国天然气的家庭使用量做出较为精准的预测。

考虑到各个市县都在加快步伐迈向天然气时代，对天然气发展有一定的政策鼓励，这样的发展政策会使得天然气管道铺设进度相对之前的模型有一个更高速度的提升。假设计划从 2021 年开始的未来五年内全国的管道铺设量每年都比现在所预估的增速翻倍，重新设计数学模型，分析并预测 2025 年全国的天然气的家庭使用量。

随着城镇天然气供应量的增大，城市燃气管道安全问题日益被关注。在城市家庭住户中的燃气泄漏更是成为人们关注的焦点。现在收集了一个无风厨房空间的灶具燃气泄漏点处的垂直轴面上的扩散气体浓度数据，建立合适的数学模型进行数值模拟验证，同时请对灶具正对面厨房门口的浓度值进行预测。假定门口正上方有 1.9 米处有燃气报警装置，报警器检测到附近的天然气浓度达到 0.5% 就会响起警铃，试问大概多少时间后会触发警铃。在有窗户且有风（风速为 1.5m/s）的情况下，结果又将发生什么变化？

## 二、 问题分析

### 2.1 问题一分析

在问题一中、我们使用了主成分分析法和灰色预测法进行预测。由于我们难以确定各个因素对于燃气普及率的影响，我们可以使用灰色系统预测的  $GM(1,1)$  模型 [1] 来预测城镇的燃气普及率。在分析人工煤气的变化趋势时，由于我们的数据包含不易确定具体作用的多个维度，我们可以使用主成分分析的方法进行数据降维，从而更方便的分析天然气的家庭使用量。

### 2.2 问题二、三分析

由于问题三中使用的模型是问题二模型的变形，我们可以把这两个问题一起分析。在问题二中，我们需要在已知的数据的基础上，确定各个因素与天然气家庭用量的相关性，并在此基础上构建天然气家庭使用量连续增长模型，使用这一模型对未来的全国天然气的家庭使用量进行预测。我们首先使用 Pearson 相关系数和偏相关系数对数据中的各个因素进行相关性分析，确定了天然气管道长度与人工煤气的家庭用量这两个对于天然气家庭用量影响最大的因素，并在此基础上通过岭回归方法确认了天然气管道长度与人工煤气的家庭用量对天然气家庭用量的影响关系。之后，为了确定天然气家庭用量与时间的关系，我们通过线性回归确定了天然气管道长度与人工煤气的家庭用量随时间的变化趋势，最终综合上面的两个模型实现了对于未来的天然气家庭用量的预测。之后，我们根据变化率的改变合理的调整了天然气管道长度与时间的关系，并且使用与问题二相似的方法完成了对问题三的预测。

### 2.3 问题四分析

首先，我们可以根据给出的数据进行绘图，并与常见的曲面进行对比，确定能较好符合数据的函数模型。我们发现椭圆抛物面模型可以较好的符合数据中的天然气质量密度在空间中的分布情况。之后可以依据椭圆抛物面方程的形式，建立时间与椭圆抛物面方程的关系，来确定不同时空分布下对应的天然气质量密度，从而确定报警器这一特定空间位置天然气浓度达到一定阈值的时间。

## 三、 模型假设与符号说明

### 3.1 模型假设

- 在预测天然气家庭用量的发展趋势时，忽略过程中可能存在的地区政策等因素带来的影响

- 在分析天然气扩散时，忽略障碍物对于扩散的影响

### 3.2 符号说明

表 1 符号说明

符号	含义
$x^{(1)}(k), y(k)$	第 $k$ 年的燃气普及率
$x^{(0)}(k)$	第 $k$ 年相对于第 $k - 1$ 年增长的燃气普及率
$r$	相关系数
$r_p$	相关系数
$U_{CG}$	表示煤气的家庭用量
$P_{NG}$	表示天然气的管道长度
$p$	天然气的泄漏流量

## 四、模型一：关于燃气发展状况的预测模型

### 4.1 对燃气普及率的预测模型

对于问题一的第一小问，需要判断燃气普及率最早在哪一年实现百分之百。我们观察了 2006 年（使用新方法计算燃气普及率元年）至 2018 年燃气普及率的每年增长情况，建立灰色预测模型 GM(1,1)。

灰色预测通过鉴别系统因素之间发展趋势的相异程度，生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来的发展趋势。我们把每年的燃气普及率和上一年的燃气普及率做差，得到每年相对于上一年燃气普及率增长的量，这些增长量符合灰色预测的特点。

以 2006 年为纪年元年，令  $x^{(1)}(k)$  表示第  $k$  年的燃气普及率，令  $x^{(0)}(k)$  表示第  $k$  年相对于第  $k - 1$  年增长的燃气普及率，即：

$$x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k - 1) \quad (1)$$

对  $x^{(1)}(k)$  建立一阶线性微分方程：

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (2)$$

其中  $a, u$  为待定系数, 分别称为发展系数和灰色作用量。

因为:

$$\frac{dx^{(1)}}{dt} = x^{(1)}(k) - x^{(1)}(k-1) = x^{(0)}(k) \quad (3)$$

得:

$$x^{(0)}(k) + ax^{(1)} = u \quad (4)$$

化简得到:

$$x^{(0)}(k) = \frac{u - ax^{(1)}(k-1)}{1+a} \quad (5)$$

使用最小二乘法估计参数  $a, u$ , 令:

$$\hat{a} = \begin{pmatrix} a \\ u \end{pmatrix} \quad (6)$$

$$B = \begin{pmatrix} -(\alpha x^{(1)}(1) + (1-\alpha)x^{(1)}(2)) & 1 \\ -(\alpha x^{(1)}(2) + (1-\alpha)x^{(1)}(3)) & 1 \\ \vdots & \\ -(\alpha x^{(1)}(n-1) + (1-\alpha)x^{(1)}(n)) & 1 \end{pmatrix} \quad (7)$$

$$Y = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T \quad (8)$$

有:

$$Y = B\hat{a} \quad (9)$$

$\hat{a}$  的一个最小二乘解为:

$$\hat{a} = (B^T B)^{-1} B^T Y \quad (10)$$

解得

$$a = 0.185 \quad b = 3.9 \quad (11)$$

将所求得的系数代入原方程中, 递推求得 2018 年之后 50 年的普及率增长量  $x^{(0)}(k)$ ,

令

$$y(k) = x^{(1)}(k) + 79.11 = \sum_{t=1}^k x^{(0)}(t) + 79.11 \quad (12)$$

其中 79.11 为 2006 年采用新的计算方法求得的燃气普及率。

则  $y(k)$  大于 100 的第一年为普及率达到 100% 的年份。

#### 4.2 关于人工煤气的主成分分析模型

第二小问要求人工煤气在家庭使用中消失的时间,我们取变量  $X = [X_1, X_2, X_3, X_4]$ , 其中  $X_1 \sim X_4$  分别表示人工煤气的供气总量、家庭用量、用气人口和管道长度。我们着重研究这四个变量下降的趋势, 即从 2003 年开始取值。考虑到数据的维度较大(四维)并且数据具有相似性, 我们使用主成分分析的方法对数据进行降维。

假设一组相关变量  $[X_1, X_2, \dots, X_n]$ , 并且是平均偏差的形式(减去均值), 主成分分析的目的是找到一个正交矩阵  $A = [a_1, a_2, \dots, a_n]$  确定的线性变换  $X = AY$ , 或:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [a_1, a_2, \dots, a_n] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad (13)$$

使得  $Y$  的变量之间两两不相关, 即  $Y$  的协方差矩阵为对角矩阵。

假设  $X$  的协方差矩阵为  $S$ , 可以证明  $Y$  的协方差矩阵为  $A^T S A$ , 因为  $S$  是对称矩阵, 所以  $S$  可以对角化:

$$S = P D P^T \quad (14)$$

其中  $P$  为  $S$  的特征向量组成的正交矩阵,  $D$  是以  $S$  的特征值为对角线的对角矩阵。当我们取  $A = P$  时,  $Y$  的协方差矩阵为

$$P^T S P = D \quad (15)$$

说明当线性变换的正交矩阵取为  $P$  时, 可以使得  $Y$  的协方差为对角矩阵。

在矩阵  $D$  中, 把  $S$  的特征值进行排列, 从左上角到右下角递减, 假设  $\sigma_1, \sigma_2, \dots, \sigma_n$  分别为  $S$  的  $n$  个特征值, 有

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n \end{pmatrix} \quad (16)$$

且有:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \quad (17)$$

协方差矩阵的特征向量  $u_1, u_2, \dots, u_3$  称为数据的主成分，第一主成分是最大的特征值对应的特征向量，以此类推。

令：

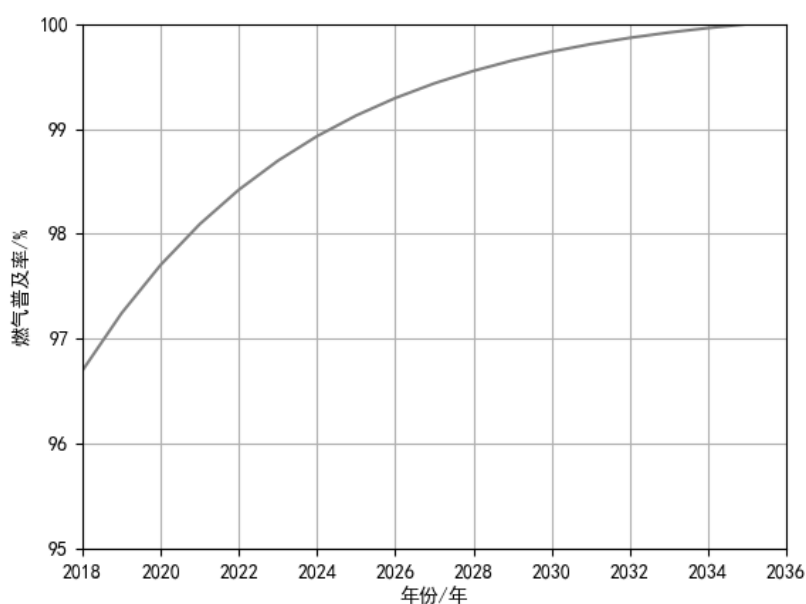
$$p_i = \frac{\sigma_i}{\sum_{j=1}^n \sigma_j} \quad (18)$$

$p_i$  称为第  $i$  主成分的贡献率，显然贡献率是递减的，我们只取前面几个主成分，使其贡献率达到一定的比例，从而达到数据降维的效果。

## 4.3 模型求解

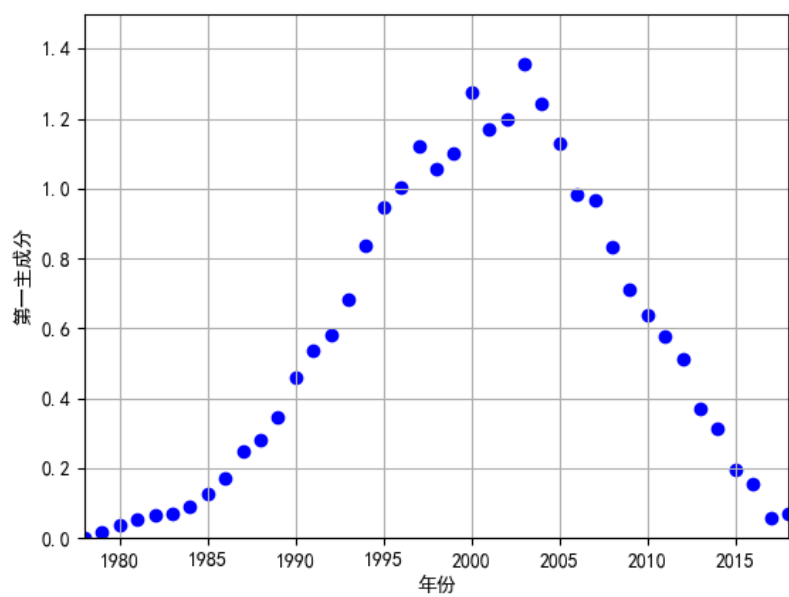
### 4.3.1 GM(1,1) 模型求解

根据我们使用 GM(1,1) 模型的预测，居民燃气化率将持续缓慢增长，并于 2035 年左右到达 100%。

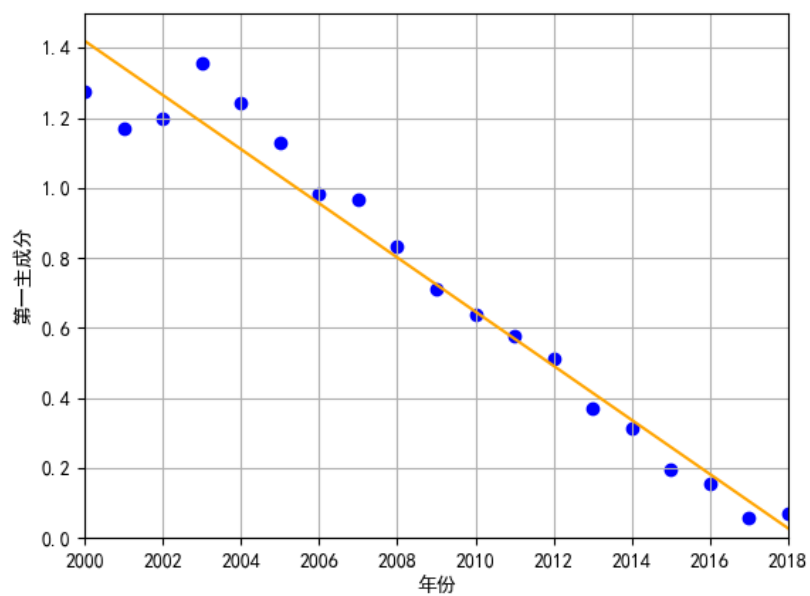


### 4.3.2 主成分分析模型求解

我们提取了两个主成分，两个主成分的贡献率分别是 (1.942569, 0.057431)。观察主成分分析的结果，我们发现在 2000-2018 年，主成分随时间的变化符合线性关系



在这种前提下，大约在 2019 年，代表煤气家庭用量的特征值将为 0，也就是家用煤气的历史将在 2019 年结束。





## 五、模型二：关于天然气家庭用量增长的回归模型

### 5.1 模型中的各个参量与天然气用量的相关性

Pearson 相关系数是最常见的相关系数，它被定义为两个变量的协方差  $Cov(X, Y)$  与两个变量的标准差  $\sigma_X, \sigma_Y$  的乘积的比值，即：

$$r(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (19)$$

由于协方差反映了衡量两个变量的总体误差，所以由协方差得到的相关系数能够在一定程度上反映两个变量的相关性。

在我国燃气产业的发展的早期，煤气产业也同样经历了一段时间的发展时期，而天然气逐步取代煤气，煤气的家用用量开始下降始于本世纪初，为了较好的反映在近年来的发展中煤气的家庭用量与天然气家庭用量间的负相关关系。我们采用煤气用量达到峰值 (2000 年) 之后的数据进行相关性分析。得到的结果如下：

表 2 参量与天然气使用量间的相关系数

参量	相关系数 $r$	参量	相关系数 $r$
煤气用气总量	-0.7358	天然气管道长度	0.9924
煤气家庭用量	-0.9318	液化石油气用气总量	-0.4941
煤气用气人口	-0.9684	液化石油气家庭用量	-0.7315
煤气管道长度	-0.9792	液化石油气用气人口	-0.7868
天然气用气总量	0.9904	液化石油气管道长度	-0.6943
天然气用气人口	0.9852		

观察上表的相关系数可以发现，天然气管道铺设长度的逐年增加而导致的天然气供气量逐年加大，以及使用人工煤气的家庭数量的逐年减少，是导致天然气家庭用量进一步提升的重要原因。

在我们使用的关于燃气产业发展的数据中，各个因素之间往往有着较大的相关性 (比如天然气的总用量影响天然气的家庭用量，而天然气的总用量受到天然气的管道长度的影响，此外天然气管道长度对于天然气的家庭用量也有直接影响)，这客观上使得我们忽略了我们关心的两个要素之间的相关关系。为了更好的表现各个变量与天然气家庭用量的相关性，我们需要使用一种能够消除其它变量影响的方法进行相关性分析。

### 5.1.1 偏相关分析

偏相关系数是在多要素所构成的系统中，当研究某一个要素对另一个要素的影响或相关程度时，把其他要素的影响视作常数，即暂时不考虑其他要素影响，单独研究两个要素之间的相互关系的密切程度时所得到的数值结果。

在分析变量  $X_1$  和  $X_2$  之间的净相关时，当控制了变量  $X_3$  的线性作用后， $X_1$  和  $X_2$  之间的一阶偏相关系数定义为：

$$r_{p12(3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} \quad (20)$$

式子中的  $r_{12}$  代表  $X_1$  和  $X_2$  的相关系数，以此类推。容易发现上式排除了  $X_1$ 、 $X_2$  和  $X_3$  的相关关系。我们依据实际情况选择了合适的控制量，并使用 Python 编程进行求解，得到的部分变量间的偏相关系数如下：

**表 3 参量间的偏相关系数**

变量 1	变量 2	控制量	偏相关系数 $r_p$
煤气用气总量	天然气家庭用量	天然气管道长度	-0.2831
煤气家庭用量	天然气家庭用量	天然气管道长度	0.4098
煤气用气人口	天然气家庭用量	天然气管道长度	0.2646
天然气用气总量	天然气家庭用量	天然气管道长度	0.1774
天然气用气人口	天然气家庭用量	天然气管道长度	-0.3260
天然气管道长度	天然气家庭用量	煤气家庭用量	0.9510
液化石油气家庭用量	天然气家庭用量	天然气管道长度	-0.0324
天然气管道长度	煤气家庭用量	天然气家庭用量	-0.6587

这表明与天然气相关的几个参数中天然气管道长度对于天然气的家庭用量有较大影响，而天然气用气总量、天然气用气人口对于天然气的家庭用量影响不大。而天然气用气人口与天然气家庭用量的偏相关系数小于 0，也反映了近年来随着生活水平的提高，人均燃气使用量增加的趋势。此外煤气的家庭用量下降对于天然气的家庭用量也有一定的影响，而其它因素对于天然气家庭用量的影响不显著。

## 5.2 回归模型的建立

根据上面对于相关系数的分析，我们可以先建立一个天然气的家庭用量关于天然气管道长度和煤气的家庭用量的回归模型来反映上述的相关性。同时建立天然气管道长度和煤气的家庭用量关于时间的变化模型，综合两个回归模型对天然气家庭用量随时间的变化趋势进行预测。

### 5.2.1 对天然气家庭用量的回归模型

我们需要，在最小二乘回归模型  $y = \beta_j x + \beta_0$  中，我们希望找到合理的  $\beta_j$  和  $\beta_0$ ，从而使得这条直线代表的预测值与真实值的距离最小。也就是使损失函数最小。

$$\sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 \quad (21)$$

根据表3，由于天然气管道长度和煤气的家庭用量可能存在一定的相关性。根据文献 [2]，我们可以采取岭回归的方法降低这种相关性对线性回归模型可能带来的影响。

岭回归相对最小二乘回归而言，在公式21对应的最小二乘回归的损失函数之上加入了一个正则项  $\lambda \sum_{j=1}^p \beta_j^2$ 。岭回归放弃了最小二乘回归的无偏性，但是对于  $\beta$  产生了一定的约束作用，在某些列之间的线性相关性比较大时采用岭回归可能能够在一定程度上减小误差。

使用岭回归和最小二乘回归得到的系数有一定的区别，用  $U_{CG}$  表示煤气的家庭用量， $P_{NG}$  表示天然气的管道长度， $C$  代表常数项，对数据进行标准化后，通过最小二乘回归得到的参数为：

$$[U_{CG}, P_{NG}, C] = [0.1584, 1.1410, 2.7756 \times 10^{-17}]$$

通过岭回归得到的参数为

$$[U_{CG}, P_{NG}, C] = [-0.1702, 0.7819, 2.7756 \times 10^{-17}]$$

根据之前的分析，由于天然气家庭用量与煤气的家庭用量负相关，与天然气管道长度正相关，这符合根据岭回归方法得到的系数。我们应该使用岭回归的结果进行分析。

### 5.2.2 结果分析与检验

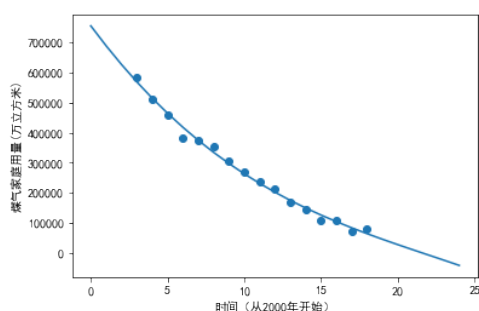
我们还可以通过把上述结果与不通过相关系数辅助分析，直接使用逐步回归方法得到的模型对比。来验证上述系数的结果。逐步回归是一种能自动从可供选择的变量中选取最重要的变量，建立回归分析模型的方法。基本思想是：将自变量逐个引入，引入的条件是其偏回归平方和经检验后是显著的。同时，每引入一个新的自变量后，要对旧

的自变量逐个检验，剔除偏回归平方和不显著的自变量。这样一直边引入边剔除，直到既无新变量引入也无旧变量删除为止。

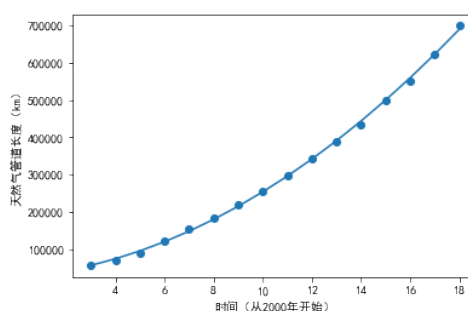
在从 2000 年开始的数据上使用逐步回归方法进行预测，最终选择的变量为 NG 管道长度和 CG 家庭用量，这与我们通过相关性分析得到的结果一致。对应的参数为  $[5.1647, 0.9550, -435901.2848]$  这我们的最小二乘回归模型的预测结果，这是因为逐步回归方法没有像岭回归一样进行正则化而导致的。

### 5.2.3 关于时间的回归模型

在上文我们已经确定了天然气的家庭用量与煤气的家庭用量以及天然气的管道长度的关系，为了确定天然气家庭用量和时间的关系，我们还需要对煤气的家庭用量以及天然气的管道长度与时间的关系进行确定。这里我们使用二次函数和指数函数的复合模型  $y = ax^2 + bx + c + d * e^{-f*x}$  对上述关系进行拟合。这里的得到的结果如下：



(a) 煤气用量与时间的关系



(b) 天然气管道长度与时间的关系

两个模型的参数值为

$$(a_1, b_1, c_1) = (12322, 2142078, 243290554, -242536296, -0.009114)$$

$$(a_2, b_2, c_2) = (1747.7, 5367.1, 25858.5, 0.3, -0.6)$$

。

上述模型的  $R^2$  值分别是 0.992 和 0.994，模型有着较好的预测效果。

### 5.3 模型应用

我们可以使用得到的回归模型可以对未来的居民天然气增长情况进行预测。对于煤气用量，根据我们的模型，煤气用量大约在 2022 年归零。所以在 2025，2030 和 2035 年对应的煤气居民用量都为 0。根据模型，在 2025 年、2030 年和 2035 年时，对应的天然气管道长度为 1260703km, 1776354km 和 2381964km。

可以根据以上信息推断得到 2025 年的天然气用量约为 4523376 万立方米, 2030 年的居民天然气用量约为 6373522 万立方米, 2035 年的居民天然气用量约为 8546438 万立方米。

#### 5.4 调整增长速率的增长模型

上文叙述的增长模型的形式, 我们可以通过求导的方式求解得到各个年份的变化率, 并通过改变变化率来反映政策的修改对于家庭天然气使用情况的影响。由于反映天然气管道长度与时间的关系的模型中的指数项的系数很小, 且年份的数值相对管道长度也明显较小, 为了简化计算, 我们可以忽略不计最后的指数项, 从而把这一模型简化为二次函数模型。

我们可以求解得到 2020 年的煤气居民用量为 28580 万立方米, 天然气管道长度为 835012km, 这一过程中没有外界政策影响。对于 2021-2025 年, 我们按照上文的算法对上面的模型进行合理简化, 得到此时对应的天然气管道长度为 1718242km, 煤气用量为 0。此时对应的天然气用量为 6165018 万立方米。

### 六、问题四：仿真模型的建立与求解

对任意的天然气泄漏流量  $p$ , 每一时刻燃气泄漏的分布情况都可由距地面 1.9m 的水平面上  $D$  关于横纵坐标的分布函数描述, 因此, 当天然气泄漏流量  $p = 0.0090m^3$  时燃气泄漏的动态分布情况可用不同时刻该水平面上天然气质量密度  $D$  关于横纵坐标以及时间  $t$  的函数进行描述。

计算报警器报警时间的关键在于利用天然气质量密度  $D$  关于横纵坐标以及时间  $t$  的函数, 求解燃气报警器所在位置天然气质量浓度等于  $3.5 \times 10^{-3}kg/m^3$  时对应的时间  $t_a$ 。

#### 6.1 三维坐标系的构建及目标点的确定

首先, 根据厨房三维坐标系示意图

设煤气泄漏点为坐标原点, 单位长度为 1m,  $x$ 、 $y$ 、 $z$  轴正方向与图2中一致, 计算得到燃气报警器所在坐标为 (3, 0, 1.1), 距离地面 1.9m 的水平面对应的平面方程为  $z = 1.1$ 。

#### 6.2 模型建立

根据附件 2 Data02.xlsx 中数据, 编写 Matlab 代码绘制得到天然气泄漏流量  $p = 0.0056m^3$  时, 燃气报警器所在水平面  $z = 1.1$  上天然气质量密度  $D$  的分布图像, 自上而下分别对应  $t = 120s, t = 60s, t = 30s$ :

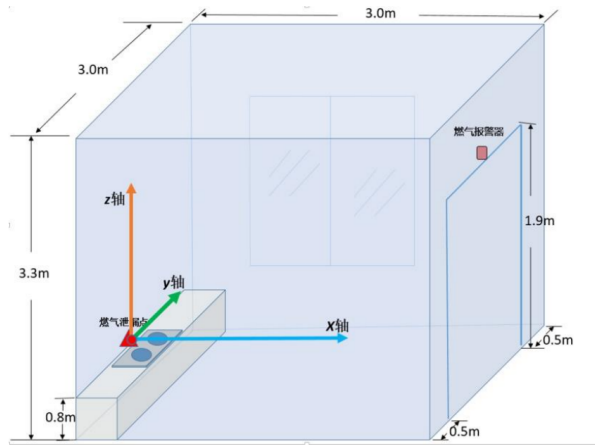


图2 厨房三维坐标系示意图

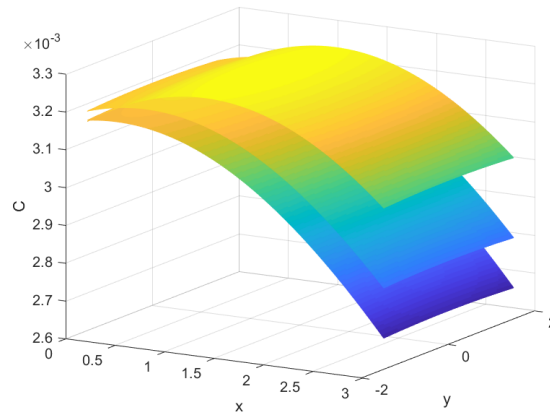


图3 平面  $z = 1.1$  上的天然气质量密度  $D$  的分布

#### 6.2.1 求解单一时刻天然气质量密度 $D$ 的分布的非线性回归模型

通过观察图3中的分布情况，可判断在  $t$  时刻，水平面  $z = 1.1$  上天然气质量密度  $D$  的函数呈椭圆抛物面方程形式，因此选择构建非线性回归模型。

设椭圆抛物面方程为

$$D = ax^2 + bx + cy^2 + d \quad (22)$$

其中  $a, b, c, d$  为待定系数。

利用最小二乘法估计待定系数  $a, b, c, d$ ，求解椭圆抛物面的方程。

#### 6.2.2 求解质量密度函数的参数与时间 $t$ 的关系的非线性回归模型

首先将椭圆抛物面方程改写为：

$$D = a'[(x - b')^2 + c'y^2] + d' \quad (23)$$

在求出时刻  $t = 30s, t = 60s, t = 120s$  分别对应椭圆抛物面方程的参数  $(a_{t=30s}, b_{t=30s}, c_{t=30s}, d_{t=30s})$ 、 $(a_{t=60s}, b_{t=60s}, c_{t=60s}, d_{t=60s})$  以及  $(a_{t=120s}, b_{t=120s}, c_{t=120s}, d_{t=120s})$  后, 根据修改后的形式求出参数  $(a'_{t=30s}, b'_{t=30s}, c'_{t=30s}, d'_{t=30s})$ 、 $(a'_{t=60s}, b'_{t=60s}, c'_{t=60s}, d'_{t=60s})$  以及  $(a'_{t=120s}, b'_{t=120s}, c'_{t=120s}, d'_{t=120s})$ 。

根据图3中的分布图像的极值点、椭圆抛物面开口大小随时间  $t$  的变化情况, 可假设  $a', b', c', d'$  分别关于  $t$  的二次函数。根据求出的参数  $(a'_{t=30s}, b'_{t=30s}, c'_{t=30s}, d'_{t=30s})$ 、 $(a'_{t=60s}, b'_{t=60s}, c'_{t=60s}, d'_{t=60s})$  以及  $(a'_{t=120s}, b'_{t=120s}, c'_{t=120s}, d'_{t=120s})$ , 构建非线性回归模型, 利用最小二乘法求解参数, 最后得到参数与时刻  $t$  的关系。

### 6.2.3 不同泄漏流量 $p$ 下天然气质量密度 $D$ 的函数的推导

根据气体扩散泄漏模型 [3], 可以得到在其他条件相同的情况下, 天然气质量密度  $D$  与天然气泄漏流量  $p$  成正比。

设初始天然气质量密度函数为  $D_0$ , 比例系数  $k = \frac{p_1}{p_0}$ ,  $p_0$  为初始天然气泄漏流量、 $p_1$  为改变后的天然气泄漏流量, 则修改天然气泄漏流量后的天然气质量密度函数  $D_1 = kD_0$ 。

### 6.3 模型求解

对于三维曲面的最小二乘拟合, 可以设总误差为

$$\delta = \sum_{i=1}^N \sum_{j=1}^N [D_{ij} - (ax_{ij}^2 + bx_{ij} + cy_{ij}^2 + d)]^2 \quad (24)$$

当  $\delta$  最小时,  $\delta$  关于系数  $a, b, c, d$  的偏导数为 0, 即

$$\frac{\partial \delta}{\partial a} = 2 \sum_{i=1}^N \sum_{j=1}^N x_{ij}^2 [D_{ij} - (ax_{ij}^2 + bx_{ij} + cy_{ij}^2 + d)] = 0 \quad (25)$$

$$\frac{\partial \delta}{\partial b} = 2 \sum_{i=1}^N \sum_{j=1}^N x_{ij} [D_{ij} - (ax_{ij}^2 + bx_{ij} + cy_{ij}^2 + d)] = 0 \quad (26)$$

$$\frac{\partial \delta}{\partial c} = 2 \sum_{i=1}^N \sum_{j=1}^N y_{ij}^2 [D_{ij} - (ax_{ij}^2 + bx_{ij} + cy_{ij}^2 + d)] = 0 \quad (27)$$

$$\frac{\partial \delta}{\partial d} = 2 \sum_{i=1}^N \sum_{j=1}^N [D_{ij} - (ax_{ij}^2 + bx_{ij} + cy_{ij}^2 + d)] = 0 \quad (28)$$

假设有矩阵

$$X = \begin{bmatrix} \sum_{i=1}^N \sum_{j=1}^N x_{ij}^4 & \sum_{i=1}^N \sum_{j=1}^N x_{ij}^3 & \sum_{i=1}^N \sum_{j=1}^N x_{ij}^2 y_{ij}^2 & \sum_{i=1}^N \sum_{j=1}^N x_{ij}^2 \\ \sum_{i=1}^N \sum_{j=1}^N x_{ij}^3 & \sum_{i=1}^N \sum_{j=1}^N x_{ij}^2 & \sum_{i=1}^N \sum_{j=1}^N x_{ij} y_{ij}^2 & \sum_{i=1}^N \sum_{j=1}^N x_{ij} \\ \sum_{i=1}^N \sum_{j=1}^N x_{ij}^2 y_{ij}^2 & \sum_{i=1}^N \sum_{j=1}^N x_{ij} y_{ij}^2 & \sum_{i=1}^N \sum_{j=1}^N y_{ij}^3 & \sum_{i=1}^N \sum_{j=1}^N y_{ij}^2 \\ \sum_{i=1}^N \sum_{j=1}^N x_{ij}^2 & \sum_{i=1}^N \sum_{j=1}^N x_{ij} & \sum_{i=1}^N \sum_{j=1}^N y_{ij}^2 & \sum_{i=1}^N \sum_{j=1}^N 1 \end{bmatrix}$$

以及向量

$$A = \begin{bmatrix} a & b & c & d \end{bmatrix}^T$$

$$Dens = \begin{bmatrix} x_{ij}^2 D_{ij} & x_{ij} D_{ij} & y_{ij}^2 D_{ij} & D_{ij} \end{bmatrix}^T$$

通过求解方程  $A = X^{-1} Dens$  即可求出方程25,26,27,28的解  $a, b, c, d$ 。

将附件 2 Data02.xlsx 中每一张表（分别为  $t = 30s$ 、 $t = 60s$  以及  $t = 120s$ ）的数据及其对应的横纵坐标分别代入到  $X, Dens$  中，解得  $(a_{t=30s}, b_{t=30s}, c_{t=30s}, d_{t=30s})$ 、 $(a_{t=60s}, b_{t=60s}, c_{t=60s}, d_{t=60s})$  以及  $(a_{t=120s}, b_{t=120s}, c_{t=120s}, d_{t=120s})$  后，求出改写后的天然气质量密度函数  $D$  中的参数

$$\begin{cases} a'_{t=30s} = -0.000064 \\ b'_{t=30s} = 0.3 \\ c'_{t=30s} = 0.0625 \\ d'_{t=30s} = 0.003169456 \end{cases} \quad \begin{cases} a'_{t=60s} = -0.000064 \\ b'_{t=60s} = 0.6 \\ c'_{t=60s} = 0.0625 \\ d'_{t=60s} = 0.003213219 \end{cases} \quad \begin{cases} a'_{t=120s} = -0.000064 \\ b'_{t=120s} = 1.2 \\ c'_{t=120s} = 0.0625 \\ d'_{t=120s} = 0.003263974 \end{cases}$$

根据得到的结果，可以将  $a'$  和  $c'$  看做与时刻  $t$  无关的常量，因此只需对  $b', d'$  进行讨论。

设  $b', d'$  关于  $t$  的函数为

$$b' = r_b t^2 + l_b t + m_b \quad (29)$$

$$d' = r_d t^2 + l_d t + m_d \quad (30)$$

对其采用最小二乘法，求解得到参数为

$$\begin{cases} r_b = 0 \\ l_b = 0.01 \\ m_b = 0 \end{cases} \quad \begin{cases} r_d = -6.80944 \times 10^{-9} \\ l_d = 2.07162 \times 10^{-6} \\ m_d = 0.00311343 \end{cases}$$



综上可得天然气泄漏流量为  $p_0 = 0.0056m^3$  时, 在水平面  $z = 1.1$  上的天然气质量密度函数为

$$D_0(x, y, t) = -0.000064[(x-0.01t)^2+0.0625y^2]-6.8094 \times 10^{-9}t^2+2.0716 \times 10^{-6}t+0.003113 \quad (31)$$

根据不同泄漏流量  $p$  下天然气质量密度  $D$  的函数的推导方法处理公式31, 得到在天然气泄漏流量为  $p_1 = 0.0090m^3$  的情况下, 在水平面  $z = 1.1$  上的天然气质量密度函数为

$$\begin{aligned} D_1(x, y, t) &= \frac{p_1}{p_0} D_0 \\ &= -0.0001029[(x - 0.01t)^2 + 0.0625y^2] - 6.8094 \times 10^{-9}t^2 + 2.0716 \times 10^{-6}t + 0.003113 \end{aligned} \quad (32)$$

即天然气泄漏流量  $p = 0.0090m^3$  时燃气泄漏的动态分布情况。

将燃气报警器坐标  $(3, 0, 1.1)$  代入公式32, 求出  $D_1 = 3.5 \times 10^{-3}kg/m^3$  对应时刻为  $t_a = 52.64s$ , 即报警器拉响警钟的时刻。

## 七、模型评价

### 7.1 模型优点

- 模型综合使用了灰色预测法, 主成分分析法等方法, 对我国燃气产业的发展进行了合理化预测
- 模型在预测中使用的模型较好的分析了各个参量与天然气家庭用量的相关性, 并且选择了岭回归等方法来降低相关性造成在预测中的不良影响
- 模型能够对扩散过程进行模拟并估计不同时间段的不同空间位置对应的天然气浓度

### 7.2 模型缺点

- 在对燃气化率预测时使用的灰色预测模型具有一定的主观性
- 模型在模拟扩散过程时没有考虑到风等外界环境因素对于模型的影响, 可以把建立的模型与高斯烟羽模型联立来增强模型的适用范围

## 参考文献

- [1] 徐宁, 丁松, 公彦德. 灰色 GM(1,1) 预测模型及拓展研究进展 [J/OL]. 数学的实践与认识:1-8

- [2] Marquardt D W, Snee R D. Ridge regression in practice[J]. The American Statistician, 1975, 29(1): 3-20.
- [3] 史宝军, 聂士明, 孙静. 基于高斯烟羽混合模型的 LPG 泄漏扩散研究 [J/OL]. 安全与环境学报:1-10[2021-07-08].<https://doi.org/10.13637/j.issn.1009-6094.2020.1517>.
- [4] Tikhonov A N. On the solution of ill-posed problems and the method of regularization[C]//Doklady Akademii Nauk. Russian Academy of Sciences, 1963, 151(3): 501-504.

## 附录 A 主成分分析——Python 源代码

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import leastsq

def func(p,x):
    return p[0]*x+p[1]

def err(p,x,y):
    return func(p,x)-y

dataf=pd.read_csv('datang.csv')
datastd=(dataf - dataf.min())/(dataf.max()-dataf.min())
datastd=datastd[['DC', 'PAG']]
datacor=datastd.corr()

print(datacor)
eig_value,eig_vector=np.linalg.eig(datacor)
eig=pd.DataFrame({"eig_value":eig_value})
eig["eig_cum"] = (eig["eig_value"]/eig["eig_value"].sum()).cumsum()
eig=eig.merge(pd.DataFrame(eig_vector).T, left_index=True, right_index=True)

loading = eig.iloc[:3,2:].T
loading["vars"]=datastd.columns

score = pd.DataFrame(np.dot(datastd,loading.iloc[:,0:1]))

print(eig['eig_cum'])
print(eig['eig_value'])
print(loading)
print(score)

year=np.array([1978,1979,1980,1981,1982,1983,1984,1985,1986,1987,1988,1989,1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018])
scoarr=np.squeeze(np.array(score).reshape((1,-1)))
print(scoarr)
lis=leastsq(err,[1,1],(year,scoarr))
print(lis)

plt.rcParams['font.sans-serif'] = ['SimHei']
plt.scatter(year,scoarr,c='blue')
#plt.plot(year,func(lis[0].tolist(),year),color='orange')
plt.xlim([1978,2018])
plt.ylim([0,1.5])
plt.ylabel('第一主成分')
```

```
plt.xlabel('年份')
plt.grid()
plt.show()
```

## 附录 B 灰色预测——Python 源代码

```
import matplotlib.pyplot as plt
import numpy as np
a = 0.185
u = 3.9
x0 = []
x1 = []
for i in range (1, 3000):
    x0.append(0)
    x1.append(0)
x1[2018] = 17.59
x=[]
y=[]
for k in range (2019, 2100):
    x0[k] = (u-a*x1[k-1])/(a+1)
    x1[k] = x1[k-1] + x0[k]
for k in range (2018, 2100):
    x.append(k)
    y.append(x1[k]+79.11)
    print(x1[k]+79.11)

plt.rcParams['font.sans-serif'] = ['SimHei']
plt.plot(x,y,color='grey')
plt.xlim([2018,2036])
plt.ylim([95,100])
plt.xlabel('年份/年')
plt.ylabel('燃气普及率/%')
plt.grid()
plt.show()
```

## 附录 C 相关性分析—Python 源代码

```
import numpy as np
import pandas as pd
from scipy.stats import pearsonr
import matplotlib.pyplot as plt
from sklearn.preprocessing import scale
```

```

from scipy.optimize import curve_fit

def partial_corr(data,a,b,c):
    r_ab=pearsonr(data[a],data[b])[0]
    r_ac=pearsonr(data[a],data[c])[0]
    r_bc=pearsonr(data[b],data[c])[0]
    r_ab_c=(r_ab-r_ac*r_bc)/(((1-r_ac**2)**0.5)*((1-r_bc**2)**0.5))
    return r_ab_c
data=pd.read_excel("Data01.xlsx")
data=data.loc[23:]
print("r=",pearsonr(data["PG用气总量"],data["NG家庭用量"]))
print("rp=",partial_corr(data,"NG管道长度","CG家庭用量","NG家庭用量"))

```

## 附录 D 逐步回归-Python 源代码

```

import numpy as np
import pandas as pd
import statsmodels.formula.api as smf
from scipy.stats import pearsonr
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.preprocessing import scale
from scipy.optimize import curve_fit

def forward_selected(data, response):
    """
    前向逐步回归算法
    使用R-squared来评判新加的参数是否提高回归中的统计显著性
    """
    remaining = set(data.columns)
    remaining.remove(response)
    selected = []
    current_score, best_new_score = 0.0, 0.0
    while remaining and current_score == best_new_score:
        scores_with_candidates = []
        for candidate in remaining:
            formula = "{} ~ {} + 1".format(response,
                                           ' + '.join(selected + [candidate]))
            score = smf.ols(formula, data).fit().rsquared_adj
            scores_with_candidates.append((score, candidate))
        scores_with_candidates.sort()
        best_new_score, best_candidate = scores_with_candidates.pop()
        if current_score < best_new_score:
            remaining.remove(best_candidate)
            selected.append(best_candidate)

```

```

        current_score = best_new_score
formula = "{} ~ {} + 1".format(response,
                                ' + '.join(selected))
model = smf.ols(formula, data).fit()
return model
def step_reg():
    data=pd.read_excel("Data01.xlsx")
    data=data.loc[23:]
    model=forward_selected(data,"NG家庭用量")
    print(model.model.formula)
    print(model.params)
    print(model.rsquared_adj)
step_reg()

```

## 附录 E 回归分析–Python 源代码

```

import numpy as np
import pandas as pd
from scipy.optimize import curve_fit
from sklearn import linear_model
from sklearn.metrics import r2_score

data=pd.read_excel("Data01.xlsx")
data=data.loc[25:]
model0=linear_model.LinearRegression(fit_intercept=True)
model=linear_model.Ridge(fit_intercept=True)
x1=np.array(data["CG家庭用量"])
x2=np.array(data["NG管道长度"])
y=np.array(data["NG家庭用量"])
data1=np.array([x1,x2]).T
xmean=x2.mean()
std_x = StandardScaler()
data1 = std_x.fit_transform(data1)
# 目标值进行标准化
mean=y.mean()
std_y = StandardScaler()
y = std_y.fit_transform(y.reshape(-1,1))
# 进行拟合
model0.fit(data1,y)
model.fit(data1,y)
print(model.coef_,model.intercept_)
print(model0.coef_,model0.intercept_)

#拟合与时间的关系
def squarex(x,a,b,c,d,e):

```

```

    return a*x**2+b*x+c+d*np.exp(-e*x)
popt,pcov=curve_fit(squarex,date,x1,maxfev=1000000)
plt.scatter(date,x1)
date3=np.arange(25)

plt.rcParams['font.family']=['SimHei']
plt.rcParams['axes.unicode_minus']=False
plt.xlabel("时间 (从2000年开始) ")
plt.ylabel("煤气家庭用量(万立方米)")
plt.plot(date3,squarex(date3,*popt))

plt.savefig("CG_fig.png")
plt.cla()
print("CG:",popt)
print("r2 CG:",r2_score(x1,squarex(date,*popt)))
print(squarex(20,*popt),squarex(25,*popt),squarex(30,*popt),squarex(35,*popt))
popt2,pcov2=curve_fit(squarex,date,x2,maxfev=1000000)
plt.scatter(date,x2)
plt.xlabel("时间 (从2000年开始) ")
plt.ylabel("天然气管道长度 (km) ")
plt.plot(date,squarex(date,*popt2))
plt.savefig("NG_fig.png")
print("NG:",popt2)
print("r2 NG:",r2_score(x2,squarex(date,*popt2)))
print(squarex(20,*popt2),squarex(25,*popt2),squarex(30,*popt2),squarex(35,*popt2))

```

## 附录 F 扩散模拟模型—matlab 源代码

```

clc;
[x,~]=meshgrid(0.00:0.01:3.00);
[~,y]=meshgrid(-1.50:0.01:1.50);
[z1,~]=xlsread('Data02.xlsx',1);
[z2,~]=xlsread('Data02.xlsx',2);
[z3,~]=xlsread('Data02.xlsx',3);
%mesh(x,y,z1),xlabel('x'),ylabel('y'),zlabel('C')
%hold on;
%mesh(x,y,z2),xlabel('x'),ylabel('y'),zlabel('C')
%hold on;
mesh(x,y,z3),xlabel('x'),ylabel('y'),zlabel('C')
hold on;
v=0.001;
vy=0;
kxy=0.22;
kz=0.2;
q=4.2896;

```

```

H=0.8;
z=1.9;
sigx = kxy * x ./ (1+0.0001*x).^0.5);
sigy = kxy * abs(y) ./ (1+0.0001*abs(y)).^0.5);
sigz = kz * z;
sqpi = sqrt(2*pi());
conf = q./(sqpi.*sqpi.*sqpi.*sigx.*sigy.*sigz);
z_ = exp(-(z - H).^2 ./ (2 * sigz.^2))+exp(-(z + H).^2 ./ (2 * sigz.^2));
t=30;
sigma1 = exp(-(x-v*(t-1)).^2./(2 * sigx.^2)).*z_.*exp(-(y-vy*(t-1)).^2 ./ (2 * sigy.^2));
for i=2:1:t
    y_ = exp(-(y-vy*(t-i)).^2 ./ (2 * sigy.^2));
    x_ = exp(-(x-v*(t-i)).^2 ./ (2 * sigx.^2));
    sigma1=sigma1+x_.*y_.*z_;
end
sigma1 = sigma1.*conf;
t=60;
sigma2 = exp(-(x-v*(t-1)).^2./(2 * sigx.^2)).*z_.*exp(-(y-vy*(t-1)).^2 ./ (2 * sigy.^2));
for i=2:1:t
    y_ = exp(-(y-vy*(t-i)).^2 ./ (2 * sigy.^2));
    x_ = exp(-(x-v*(t-i)).^2./(2 * sigx.^2));
    sigma2=sigma2+x_.*y_.*z_;
end
sigma2 = sigma2.*conf;
t=120;
sigma3 = exp(-(x-v*(t-1)).^2./(2 * sigx.^2)).*z_.*exp(-(y-vy*(t-1)).^2 ./ (2 * sigy.^2));
for i=2:1:t
    y_ = exp(-(y-vy*(t-i)).^2 ./ (2 * sigy.^2));
    x_ = exp(-(x-v*(t-i)).^2./(2 * sigx.^2));
    sigma3=sigma3+x_.*y_.*z_;
end
sigma3 = sigma3.*conf;
z=-0.000032.*(2.*(x-0.3).^2+0.125.*y.^2)+0.003169456;
%z=-0.000032.*(2.*(x-0.6).^2+0.125.*y.^2)+0.00321321934151253;
%z=-0.000032.*(2.*(x-1.2).^2+0.125.*y.^2)+0.00326397442630851;
mesh(x,y,z)
%hold on;
%mesh(x,y,sigma1),xlabel('x'),ylabel('y'),zlabel('z');
%hold on;
%mesh(x,y,sigma2),xlabel('x'),ylabel('y'),zlabel('z');
%hold on;
%mesh(x,y,sigma3),xlabel('x'),ylabel('y'),zlabel('z');

```