



Igor Pro 微分方程拟合研究

贾小文, 范海英, 尹霖, 张金凤, 赵云飞, 邱成锋

(陆军军事交通学院基础部, 天津 300161)

摘要: 使用微分方程拟合数据不同于普通的数据拟合, 拟合无明确的解析表达式可以利用, 拟合结果容易发散, 拟合程序设计较为困难。以 COVID-19 疫情数据为例, 利用病毒传播 SEIR 模型, 详细介绍了使用 Igor Pro 对微分方程进行数据拟合的基本原理、程序面板的设计、拟合函数的编写、初始参数的设置等, 分析了拟合结果的物理意义及可能存在的局限性, 讨论了利用微分方程拟合数据时需要注意的问题。

关键词: 数据拟合; 微分方程; SEIR 模型; Igor Pro

中图分类号: O4-39

文献标志码: A

DOI: 10.12179/1672-4550.20200451

The Fitting Research of Differential Equation by Igor Pro

JIA Xiaowen, FAN Haiying, YIN Lin, ZHANG Jinfeng, ZHAO Yunfei, QIU Chengfeng

(General Course Department, Army Military Transportation University of PLA, Tianjin 300161, China)

Abstract: Fitting data to a differential equation is different from ordinary data fitting cases. There is no explicit analytical expression to use in fitting. The fitting results are easy to diverge and the fitting program design is more difficult. Taking COVID-19 epidemic data as an example, this paper introduces the use of Igor Pro in detail by using the SEIR model of virus transmission. The basic principle of fitting data to differential equation by Igor Pro, the design of software panel, the coding of fitting function and the setting of initial parameters are analyzed. The physical meaning of fitting results and the possible limitations are analyzed. The problems needing attention when using differential equation to fit data are discussed.

Key words: data fitting; differential equation; SEIR model; Igor Pro

数据拟合利用已知公式模型与实验数据进行对比, 以获取最佳参数值或是实验数据的演化趋势, 对未测量点的数据做出预测。与图解法比较, 数据拟合可以排除主观因素, 找出依靠直觉无法发现的客观规律。数据拟合是实验数据处理的一项重要内容。

设实验数据是 $\{(x_{1i}, \dots, x_{2i}, \dots), y_i\}$, 则拟合的任务就是寻找最吻合数据的理论曲线(面):

$$y = f(a_1, a_2, \dots, x_1, x_2, \dots) \quad (1)$$

式中, a_i 表示拟合参数。拟合参数往往具有明确的含义, 如伏安特性曲线中, 若 y 表示电压, x 表示电流, 理论模型为 $y = a_1 x$, 则 a_1 就是待测量电阻 R 。数据拟合的基本依据是最小二乘原理:

$$\varepsilon^2 = \sum_i \frac{(y_i - y)^2}{\sigma_i^2} \quad (2)$$

式中, y_i 表示 y 的第 i 个测量值, σ_i^2 表示测量值 y_i 的标准方差, 其倒数表示对应测量值在式(2)中的权值。当式(2)取值最小时, 对应的 a_i 就是最佳拟合值。

式(1)中, 若各参数 a_i 是线性的(拟合公式可表示为基函数的线性组合), 则称为线性拟合; 若 a_i 是非线性的(a_i 要么是高次幂的, 要么本身包含在函数内), 则称为非线性拟合。无论是线性拟合还是非线性拟合, 一般都要求明确给出 $y = f(a_1, a_2, \dots, x_1, x_2, \dots)$ 的表达形式, 这在大多数情况下是符合实际的, 也是最易于理解的数据拟合形式。但是也有例外, 在有些情况下 y 的表达形式并非解析表达式, 而是由微分方程(或者是积分方程、傅里叶变换、卷积等)描述, 此时 $y = f(a_1, a_2, \dots, x_1, x_2, \dots)$ 就没有明确的解析表达式,

收稿日期: 2020-09-27; 修回日期: 2020-11-23

基金项目: 陆军军事交通学院重点课程建设项目(4142ZB101)。

作者简介: 贾小文(1983-), 男, 博士, 副教授, 主要从事凝聚态物理和实验数据分析处理研究。

普通方法是无法对这样的数学模型进行拟合的^[1-5]。对于含有微分方程的数学模型拟合方法, 文献中鲜有介绍, 本文以 COVID-19 疫情数据为例, 利用 SEIR 微分方程模型对数据进行拟合, 介绍了利用 Igor Pro 通过微分方程对数据拟合的原理和方法。

1 拟合方法和原理

1.1 Igor Pro 数据拟合功能介绍

Igor Pro 是一个数据可视化和分析处理科学软件, 以程序设计为基础, 可扩展性非常强。Igor Pro 数据拟合功能完善, 可以进行线性拟合和非线性拟合。拟合输出信息丰富, 包括拟合参数值、拟合误差、残差、拟合曲线, 以及给定置信水平的拟合分布区间和测量值的分布区间等^[6-7]。

Igor Pro 对数据的拟合方式可分为:

- 1) 利用菜单 Quick Fit 进行拟合;
- 2) 利用 Curve Fitting 对话框进行拟合;

3) 调用拟合命令 (FuncFit、CurveFit) 结合程序设计进行拟合。CurveFit 命令使用内置的数学模型进行拟合, 无需指定初始参数, 使用方便。FuncFit 命令使用自定义公式或者数学模型进行拟合, 灵活性最强, 但需要指定合适的初始参数。

本文使用 FuncFit 命令对疫情数据用 SEIR 微分方程模型进行拟合。

如果拟合公式不是内置函数, 则需要编写自定义函数。普通情况下拟合函数一次只能返回一个函数值, 本文拟合对象是一组微分方程, 一次返回一个值就不行了。Igor Pro 拟合函数也可一次返回整个微分方程的解, 这是拟合微分方程的基础。

利用 Igor Pro 进行数据拟合包括以下 4 个步骤:

- 1) 构建数学公式或者数学模型;
- 2) 指定拟合参数, 并给出初始值;
- 3) 调用拟合命令完成拟合;
- 4) 通过拟合参数给出有意义的结果。

1.2 病毒传播数学模型

病毒传播的经典模型有指数模型、SI 模型、SIS 模型、SIR 模型、SEIR 模型等, S 表示易感染者 (susceptible), I 表示感染者 (infectious), R 表示移除者 (removed), E 表示携带者 (exposed)。指数模型中, 感染人数随时间指数增长, 病毒在传播初期往往具有此特征。SI 模型 (即 Logistic 模型)

中, 感染人数增加速度先快后慢, 最后达到稳定, 群体中所有个体都会被感染。SIS 模型中, 感染的个体有一定概率恢复为健康个体 (但不具有免疫性, 还可被感染), 与 SI 模型相比较, 最后感染人数和健康人数会达到平衡。SIR 模型中, 感染的个体有一定概率恢复健康 (如被治愈) 且获得免疫性, 不会再被感染, 感染人数逐渐下降并最终消失。SEIR 模型中, 易感染个体感染后并不马上发病, 而是成为携带者 (潜伏状态), 携带者有一定概率成为感染者, 感染者有一定概率成为健康个体且获得免疫性, 群体最终将没有感染者^[8-10]。COVID-19 病毒具有明显的潜伏期, 被治愈后具有抗体, 因此符合 SEIR 模型, 可用 SEIR 模型描述。根据 COVID-19 传播的实际特征, 构建如图 1 所示 SEIR 模型。

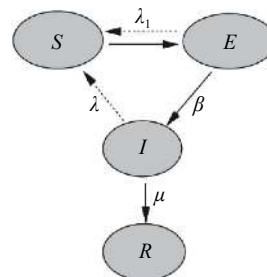


图 1 SEIR 病毒传播模型

总人数 $N=S+E+I+R$ 。易感染者 S 既有可能被感染者 I 感染, 也有可能被携带者 E 感染 (图 1 虚线所示), 被 I 感染的比例系数为 λ , 被 E 感染的比例系数为 λ_1 ; 携带者 E 有一定概率会成为感染者 I , 概率为 β ; 感染者 I 有一定概率成为移除者, 移除者包括恢复健康个体 (具有抗体, 不再被感染) 和死亡个体, 概率为 μ ; 有效再生数为 $R_0\left(\frac{\lambda}{\mu}=R_0\right)$ 。 S 、 E 、 I 、 R 满足如下微分方程组:

$$\begin{cases} \frac{dS}{dt} = -\frac{\lambda IS}{N} - \frac{\lambda_1 ES}{N} \\ \frac{dE}{dt} = \frac{\lambda IS}{N} + \frac{\lambda_1 ES}{N} - \beta E \\ \frac{dI}{dt} = \beta E - \mu I \\ \frac{dR}{dt} = \mu I \end{cases} \quad (3)$$

2 利用 SEIR 模型拟合疫情数据

2.1 数据来源

本文使用的数据为中国 COVID-19 疫情数据。由于中国政府高效的疫情防控措施, 在

2020 年 3 月上旬以后，全国各省国内每日新增确诊人数已经基本为零，新增病例主要以境外输入为主，如图 2 所示。境外输入病例被严格管制，其传播途径已被人为切断，不再满足 SEIR 模型假设。故本文数据截止日期选取为 2020 年 3 月 11 日。

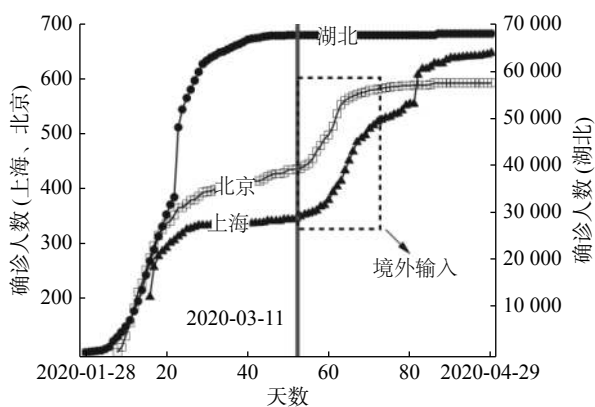


图 2 疫情数据随时间变化关系

一些互联网公司将会全国和各省市卫健委官网公布的数据汇总为实时疫情地图数据。本文从腾讯网站获取这些数据，腾讯网站提供 JSON 格式的数据，JSON 数据由纯文本组成，是一种轻量级的规范的数据格式，广泛用于互联网数据传输^[11]。利用 Igor Pro 的 FetchURL 函数可获取数据：String s0=FetchURL(url)，url 表示数据所在的网络地址。利用 JSON XOP 扩展解析和提取数据^[12]。

2.2 拟合程序设计

以每日累积确诊数、治愈数和死亡数作为原始数据。其中“每日累积确诊数-治愈数-死亡数”对应 SEIR 模型中感染人数 I ，“治愈数+死亡数”对应移除人数 R 。根据微分方程组(3)，引入 6 个拟合参数，含义如表 1 所示。

表 1 SEIR 模型拟合参数及其含义

序号	参数	含义
1	s_0	易感染人数初始值(S_0)
2	e_0	携带者人数初始值(E_0)
3	n	易感者被感染者感染的比例系数(λ)
4	n_1	易感者被携带者感染的比例系数(λ_1)
5	b	携带者成为感染者的概率(β)
6	u	感染者被治愈的概率(μ)

由于易感人群数 S 和携带人群数 E 是未知的，因此将初始值设置为拟合参数，通过拟合已有的感染者来估计易感人群和携带人群数量。通

过拟合参数 λ 和 μ 可计算有效再生数 R_0 。Igor Pro 拟合函数(FuncFit 命令会调用此函数对数据进行拟合)设计如下：

```
Function FitSEIRModel(pw,yw,xw):FitFunc
    Wave pw,yw,xw
    SVAR wIName,wRname
    Variable s0,e0,n,n1,b,u,X0
    s0=pw[0]
    e0=pw[1]
    n=pw[2]
    n1=pw[3]
    b=pw[4]
    u=pw[5]
    wave wI=$wIname //感染者数据
    wave wR=$wRname //移除者数据
    Variable N0=numpts(wI)
    Make/N=(N0,4)/O wtmp
    wtmp[0][0]=s0
    wtmp[0][1]=e0
    wtmp[0][2]=wI[0]
    wtmp[0][3]=wR[0]
    IntegrateODE SEIRDifFunc,pw,wtmp
    yw=wtmp[p][2]
End
```

拟合函数 FitSEIRModel 利用微分方程组(3)根据 pw 提供的拟合参数计算理论值，pw 参数含义如表 1 所示。被拟合数据为当前感染人数 I 。函数 FitSEIRModel 调用 IntegrateODE 求解微分方程组(3)，返回的拟合值是 yw，yw 是一个 wave，即一组数据，而非单个值。将拟合函数的第二个参数指定为一个 wave(如这里的 yw)，Igor Pro 会将该参数 yw 作为函数的返回值，因此只需要在拟合函数里对 yw 赋值就可以了。IntegrateODE 是一阶微分方程求解命令，可求解 $Y' = f(X)$ 类型的微分方程 (Y 和 X 是矢量)。SEIRDifFunc 是微分方程组(3)在 Igor Pro 中的函数表示，即：

```
Function SEIRDifFunc(pw,xx,yw,dydx)
    Wave pw
    Variable xx
    Wave yw,dydx
    Variable s0,e0,n,n1,b,u
    s0=pw[0]
    e0=pw[1]
```

```

n=pw[2]
n1=pw[3]
b=pw[4]
u=pw[5]
Variable S,E,I,R,N0
S=yw[0]
E=yw[1]
I=yw[2]
R=yw[3]
N0=S+E+I+R
dydx[0]=-n*I*S/N0-n1*E*S/N0
dydx[1]=n*I*S/N0+n1*E*S/N0-b*E
dydx[2]=b*E-u*I
dydx[3]=u*I

```

End

拟合命令为 $\text{FuncFit/L}=(N)/Q \text{ FitSEIRModel}$, FitCoef , $wI/D/C=\text{constrW}$ 。wI 为感染者数据 I, constrW 限定除了 λ_1 外其他参数不小于 0。FuncFit 命令反复调用 FitSEIRModel 函数, 利用 FitCoef 里的参数值求解微分方程组 (3), 并将求解结果 (理论值) 与疫情数据 (实验值) 比较, 直到理论值与实验值吻合为止, 此时 FitCoef 里的拟合参数值就是各参量的期望值。

由于使用微分方程拟合且参数数量较多, 必须给出较为合理的初始参数, 否则拟合很容易发散。本文先手动设置参数, 利用该参数计算 SEIR 理论值 (利用 SEIRDifFunc 函数计算) 并与实验值比较, 采用接近实验值的参数为初始参数。为了方便参数设置和数据管理, 将整个拟合过程设计为程序面板: 选择并获取 (下载) 疫情数据 (全国、各省市), 设定初始参数并观察初始参数所得理论值是否与观测值一致, 基本一致后进行拟合, 拟合的结果显示在程序窗口, 保存拟合结果。程序界面如图 3 所示。

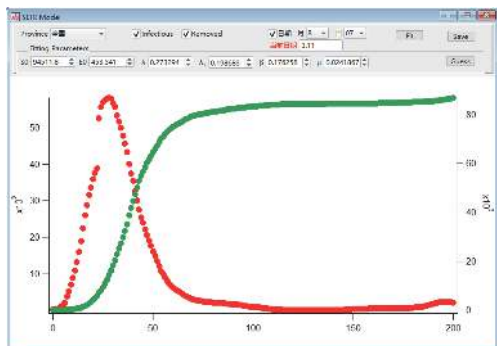


图 3 数据分析拟合程序面板

2.3 拟合结果

以全国和湖北、非湖北和广东的数据为例, 拟合结果如图 4 所示。图形标记表示真实疫情数据, 黑色实线表示拟合数据, 可以看出, SEIR 模型与实测数据吻合度较高, 说明拟合方法是有效的。

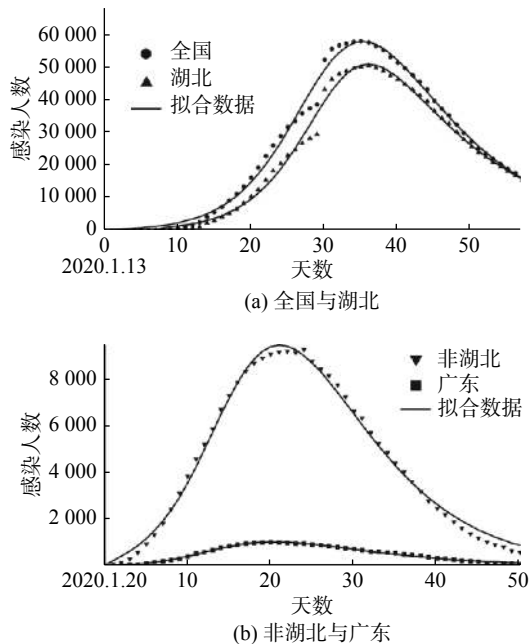


图 4 拟合结果

为了更进一步地说明拟合的可靠性, 在表 2 中给出了全国和湖北、非湖北和广东的详细拟合结果。 S_0 和 E_0 表示拟合的易感人群数和携带人群数初始值, 这里可见易感人群数并非各地区总人口数, 这可能与管控措施有关, 大量的人群由于居家隔离等原因被排除在“模型里的易感人群”之外。 λ 表示易感者被感染者感染的比例系数, 即感染者对易感人群的感染能力, 值越大表示被感染者感染的人数越多, 从表 2 可以看出湖北和全国的该参数值远大于非湖北地区, 说明湖北地区感染人数多密度高, 更容易造成对易感人群的感染。 λ_1 表示易感者被携带者感染的比例系数, 即携带者对易感人群的感染能力, 值越大表示由携带者导致的感染人数比例越大。这里比较异常的是全国和湖北该参数拟合值为负值, 非湖北和广东拟合值为正值, 说明在湖北可能存在未就医自愈的情况, 而全国其他省份的感染者则有很大比例是由携带者感染所致。 β 表示携带者转化为感染者的概率, 全国和湖北该参数值都较大, 非湖北则比较小, 这预示湖北的携带者被确诊的概率很大, 而其他省则小一些。 μ 表示移除率 (治愈率+死亡率, 主要贡献是治愈率), 非湖北地区该值明显

高于湖北地区,这可能与湖北地区相对不足的医疗资源和高密度的疫情环境有关。根据 1.2 节 SEIR 模型, $R_0 = \lambda/\mu$, 从表 2 可以看出全国和湖北的有效再生数都非常高,特别是湖北达到了 6.2,而非湖北有效再生数则非常小,全国除去湖北有效再生数约为 1.2,疫情程度仅次于湖北的广东该

值为 1.1,显著低于湖北。有效再生数再次说明了疫情在湖北严重的原因,也说明了全国其他省市采取的延迟上班、居家隔离等防控措施是非常有效的。由于全国累积确诊数据 83% 以上贡献来自于湖北,因此全国的数据行为和湖北基本一致,而非湖北的数据则和疫情第二严重的广东基本一致。

表 2 全国、非湖北、湖北和广东省拟合结果及有效再生数 R_0

数据来源	S_0	E_0	λ	λ_1	β	μ	R_0
全国	165730	631.01	0.41146	-0.21933	0.93647	0.096291	4.3
非湖北	39253	821.27	0.19201	0.37932	0.23096	0.15876	1.2
湖北	115180	1235.8	0.48267	-0.91308	1.4504	0.078183	6.2
广东	3582.2	156.01	0.15428	0.41099	0.14367	0.14002	1.1

3 结束语

SEIR 模型是一个理想模型,易感染者、携带者、感染者和移除者按照比较简单的比例关系相互转化,参数取值是固定不变的,但实际中,由于政府会根据疫情情况采取不同程度的防控措施,医疗资源的投入也会实时变化,这些都会动态改变模型中各参数取值,除此之外,还有可能存在另外的一些因素是 SEIR 模型所未能涵盖的,因此对疫情数据的准确建模是一个非常复杂的问题,本文图 4 和表 2 中的拟合结果仅仅是根据 1.2 节中的简单模型得出的结论,一些参数的物理意义及其结论从统计学和传染病流行病学原理来讲还需要进一步确认。

Igor Pro 拟合函数可以一次返回一组数据(wave),利用这个特性,当待拟合的数学模型里含有积分、傅里叶变换、卷积、微分等操作时,可以在拟合函数里对拟合公式施以相应变换,然后将变换后的结果作为拟合函数的返回值就可以了,本文中微分方程的拟合正是基于这样的原理。但是也必须注意,与一次返回一个值的普通拟合方式比较,从数学上讲,一次返回一组值更复杂,前者是将单个自变量 x 处的拟合值与实测值逐个比较,而后者则是将所有自变量 x 处的拟合值与实测值整体比较,显然前者的约束作用更强,因此拟合也更容易收敛,而后者则不太容易收敛,这是符合实际的。本文采取了手动设置参数的方式,在拟合之初,通过手动尝试不同参数,以尽可能获取接近实验数据的参数取值,然后再开始拟合,避免了拟合发散或者收敛到不合理理解处的问题,保证了拟合的成功率。

参考文献

- [1] 曾铭,王晋研,王刚志.物理实验数据的线性与非线性拟合[J].大学物理实验,2017,30(2):116-119.
- [2] 唐家德.基于MATLAB的非线性曲线拟合[J].计算机与现代化,2008(6):15-19.
- [3] 姜云,康智慧,王丛石,等.用最小二乘法对多峰值数据进行曲线拟合及曲线峰中心位置的精确求解[J].物理实验,2003(11):38-39.
- [4] 王宇,胡赞,曹仲义,等.Origin软件在化工原理实验数据拟合中的应用[J].实验技术与管理,2010,27(1):86-88.
- [5] 赵宝贵.Matlab在数据拟合中的应用[J].科技广场,2007(1):145-146.
- [6] 贾小文.Igor Pro在物理实验数据处理中的应用[J].实验室科学,2020,23(2):88-92.
- [7] 贾小文.Igor Pro实用教程[M].北京:清华大学出版社,2018.
- [8] 周涛,刘权辉,杨紫陌,等.新型冠状病毒感染肺炎基本再生数的初步预测[J/OL].中国循证医学杂志,2020,20(3):1-6[2020-02-19].<http://kns.cnki.net/kcms/detail/51.1656.r.20200204.1640.002.html>.
- [9] 耿辉,徐安定,王晓艳,等.基于SEIR模型分析相关干预措施在新型冠状病毒肺炎疫情中的作用[J/OL].暨南大学学报(自然科学与医学版):2020:1-7[2020-02-19].<http://kns.cnki.net/kcms/detail/44.1282.n.20200214.1318.002.html>.
- [10] 王志心,刘治,刘兆军.基于机器学习的新型冠状病毒(2019-nCoV)疫情分析及预测[J/OL].生物医学工程研究,2020,39(1):1-9[2020-02-19].<http://kns.cnki.net/kcms/detail/37.1413.R.20200213.0956.002.html>.
- [11] 腾讯网站.新型冠状病毒肺炎疫情实时追踪[EB/OL].[2020-03-01].<https://news.qq.com/zt2020/page/feiyuan.htm>.
- [12] BRAUN T. JSON XOP for Igor Pro[EB/OL].[2020-03-01].<https://docs.byte-physics.de/json-xop/>.2020.03.