

目 录

摘要	ii
Abstract	iii
第一章 引言	1
1.1 研究背景	1
1.2 文献综述	1
1.3 研究目标与本文结构	2
第二章 PBPK 模型的建立	3
2.1 模型的实验与生理背景	3
2.2 模型的数学形式	4
2.3 PBPK 模型的求解	9
第三章 PBPK 模型参数反演的神经网络方法: 使用固定时间节点数据	11
3.1 参数反演的背景与网络模型的基本介绍	11
3.2 参数反演神经网络模型的数据集结构与构建	12
3.2.1 网络功能与数据集结构介绍	12
3.2.2 标签集的构建	12
3.2.3 特征集的构建	14
3.3 参数反演神经网络模型的架构	17
3.4 参数反演神经网络模型的训练	18
3.4.1 数据集的预处理, 划分与训练流程	18
3.4.2 损失函数与优化器	18
3.5 对偶学习与 PBPK 模型拟合神经网络模型	19
3.5.1 对偶学习的引入背景	19
3.5.2 正向拟合神经网络模型的数据集结构与构建	19
3.5.3 正向拟合神经网络模型的网络结构与训练	20
3.6 参数反演神经网络模型的效果展示	20
第四章 PBPK 模型参数反演的神经网络方法: 使用随机时间节点数据	24
4.1 参数校准与网络模型的基本介绍	24
4.2 参数反演神经网络模型的数据集结构与构建	25
4.3 参数反演神经网络模型的网络结构与训练	26
4.4 可输入时间信息的 PBPK 模型拟合神经网络模型	26
4.5 参数反演神经网络模型的效果展示	27

4.6 基于真实实验数据 ^[1] 的参数校准	29
第五章 PBPK 模型的统计学分析	30
5.1 敏感性分析	30
5.2 不确定性分析	30
第六章 总结	32
附录 A PBPK 模型物理量名称-含义对照表	34
参考文献	39
致谢	41

摘要

本研究旨在通过建立神经网络模型来实现对双酚 S(BPS) 在人体内的生理药代动力学模型 (PBPK) 内重要参数的反演, 以更快的速度与准确性完成模型的校准。

首先, 本文介绍了 PBPK 模型的建立, 包括模型的实验与生理背景, 数学形式和求解方法。随后, 本文详细阐述了使用固定时间节点数据与使用随机时间节点数据的 PBPK 模型参数反演神经网络方法, 并通过数据集结构与构建、网络模型架构、训练流程和效果展示等环节, 展示了该方法的有效性。此外, 本文使用训练好的神经网络模型基于真实实验数据完成了 PBPK 模型的参数校准。最后, 本文评估了模型输出对校准后的皮肤三参数组的敏感度和不确定性。

本文通过数值实验证明了神经网络模型在 PBPK 模型框架下能够胜任参数反演与参数校准的任务, 并建议在使用神经网络模型作为校准工具的前提下, 研究者在人体实验中可以减少对受试者的采血采尿频次, 以降低成本。

关键词: 双酚 S(BPS); 生理药代动力学模型 (PBPK); 深度学习; 参数反演

CLC code: O24

Abstract

This study aims to establish a neural network model for the inversion of key parameters within the physiologically based pharmacokinetic (PBPK) model for bisphenol S (BPS) in the human body, to calibrate the model more rapidly and accurately.

Initially, this paper introduces the development of the PBPK model, including its experimental and physiological background, mathematical formulation, and solution methods. Subsequently, the paper elaborates on the neural network methods for PBPK model parameter inversion using both fixed and random time node data. The effectiveness of these methods is demonstrated through the construction of the dataset, the architecture of the network model, the training process, and the presentation of results. Additionally, this paper employs the trained neural network model to calibrate the PBPK model parameters based on real experimental data. Finally, the paper assesses the sensitivity and uncertainty of the model output to the calibrated skin triplet parameters.

Through numerical experiments, this paper demonstrates that the neural network model is capable of performing parameter inversion and calibration tasks within the framework of the PBPK model. It is suggested that, with the neural network model as a calibration tool, researchers can reduce the frequency of blood and urine collection from subjects in human experiments, thereby reducing costs.

Keywords: Bisphenol S (BPS); Physiologically Based Pharmacokinetic (PBPK) Model; Deep Learning; Parameter Inversion

CLC code: O24

第一章 引言

1.1 研究背景

双酚类物质 (Bisphenols, BPs) 是一种工业用化学物质, 被大量用于生产聚碳酸酯及环氧树脂^[2]. 这两种可能会含有 BPs 的高分子物质又常被投入生产食品接触材料或其他日常使用材料, 例如塑料杯, 奶瓶, 纸币, 金属涂层等^[3]. 在日常生活中, BPs 通过皮肤渗透与口服摄入两种主要途径进入人体内环境, 参与后续的分布与代谢. 双酚 A (BPA), 作为最早投入工业生产的 BPs, 已被证实对人体具有毒性^[4]. 事实上, BPA 会对人体的多个系统 (如呼吸系统, 神经系统, 生殖系统) 造成损害^[5]. BPA 与双酚 S (BPS) 两种 BPs 经口服进入人体后, 经消化系统来到小肠, 并在此分别葡萄糖醛酸化为 BPA-g 与 BPS-g, 葡萄糖醛酸化后的双酚物质会进入血液循环并最终随尿液排出体外; 未葡萄糖醛酸化的 BPA 或 BPS 将会进入肝脏并在此被部分磷酸化为 BPA-s 或 BPS-s, 部分 BPs 在肝脏仍会被葡萄糖醛酸化, 这些衍生物与未发生反应的 BPs 都会直接进入血液循环并最终随尿液排出体外^[4,6]. 同时, 在小肠或肝脏处进入血液循环的 BPs 会随血液进入人体的各个器官, 如脑, 生殖腺等. 若 BPs 经由皮肤渗透进入人体, 将会直接进入血液循环并跟随血液到达各个器官, 其中进入小肠或肝脏的部分 BPs 将会根据所处位置被葡萄糖醛酸化或是被磷酸化. 为了找寻比 BPA 更安全的替代品, 研究 BPS 等双酚物质在人体中的代谢过程是有必要的^[7].

生理药代动力学模型 (Physiology-Based Pharmacokinetic Model, PBPK) 是药学中定量描述化学品在人体中吸收, 分布, 代谢, 排泄过程的经典模型, 常被用于化学品生态风险评价, 人类健康风险评估以及药物开发^[8]. PBPK 模型将包含血浆在内的对目标化学品特异性较强的靶点组织器官抽象为一个个“房室”, 以质量守恒定律和相关生化反应为基础定量计算目标化学品在各房室之间的交换与各房室之内的代谢过程^[9]. 当某些靶点组织器官的目标化学品含量难以实际测出时, PBTK 模型的结果能够提供一个良好的预测^[8]. 只需要确定 PBPK 模型中重要参数的数值, 就能在脱离实际人体实验的情况下给出人体吸收目标化学品后靶点组织器官的化学品含量.

1.2 文献综述

Yang 等人^[10]在 2015 年首次建立了使用人类参数的 BPA 在生物体内的 PBPK 模型, 该模型基于口服摄入的吸收方式, 共设置了 10 个仓室, 分别为血浆, 肝脏, 脂肪, 性腺, 血流丰富组织, 血流缓慢组织, 大脑和皮肤, 剩下两个仓室分别是 BPA-g 和 BPA-s 的反应仓室. 为了校准模型, Yang 等人设立了成年人类口服 BPA 的人体实验并记录了受试者尿液中的 BPA 的含量信息, 借助真实数据, 研究者使用视觉拟合与传统优化算法对 PBPK 模型中与肝脏有关的参数进行了优化校准. 最后, Yang 等人利用敏感性分析与不确定性分析评估了模型. Karrer 等人^[11]在 2018 年重新调整了此 PBPK 模型, 提供了 BPA 的其他双酚类替代品的模型参数, 并增加了通过皮肤渗透吸收 BPs 的情形. Karrer 等人利用人体 BPS 接触实验的血液与尿液数据完成了对模型结构的调整, 并同样使用了视觉拟合与传统优化算法对部分参数进行了调整, 最终也使用敏感性与不确定性分析评估了模型. Hu 等人^[1]在 2023 年对皮肤渗透模型进行了改进, 在原本皮肤作为单独仓室的基础上将其分割

成了五个小仓室, 分别为表皮储仓, 角质层, 活性表皮, 毛囊以及未参与渗透吸收的未暴露皮肤. 该文章也设置了志愿者实验, 利用 BPS 暴露后受试者尿液中 BPS 与 BPS-g 的含量来优化 PBPK 模型与皮肤仓室相关的三个参数, 并使用敏感性与不确定性分析评估了修改后的 PBPK 模型.

1.3 研究目标与本文结构

Hu 等人^[4]文章的参数优化部分中使用了传统的优化算法, 且不确定性分析中的蒙特卡洛模拟需要对模型进行上万次的计算, 在计算机上运行的时间较久, 有较大的优化空间. 本文欲根据 Hu 等人在论文中提供的 PBPK 模型的数学形式来重建 PBPK 模型的求解模型, 并使用深度学习的办法完成参数校准的工作, 最终在校准结果的基础上完成模型的敏感性分析与不确定性分析.

本文包含三个主要结构, 分别为 PBPK 模型的求解, PBPK 模型的参数校准与 PBPK 模型的统计学分析. 其中使用神经网络模型完成参数校准是本文的主题与重点, 本文将使用深度残差连接网络架构训练出两种神经网络模型. 网络模型的输入为离散的 BPS 含量信息, 输出为反演得到的需校准的 PBPK 模型参数, 以达到参数校准的目的. 两种神经网络模型分别为不可输入时间信息的网络和可输入时间信息的网络, 输入时间信息意为网络在接收 BPS 含量信息的同时还能接收该含量信息所处的时间点的信息.

第二章 PBPK 模型的建立

Hu 等人^[1]在 github 中共享了论文中的数据以及部分代码^[12]. 共享中包含了 PBPK 模型和参数优化所使用的受试者真实数据等, 本章内容参照了这些工作. 该模型对多种双酚类物质都适用, 本文后续只讨论双酚 S(BPS) 的情形, 且只考虑皮肤暴露途径的外源 BPS 输入.

2.1 模型的实验与生理背景

Hu^[1]等人通过使用含有氘代 BPS(BPS-d8) 的热敏纸摩擦手指的方式令受试者暴露于 BPS. 受试者接触热敏纸共 $\frac{1}{6}h$, 脱离热敏纸后再等待 $2h$, $\frac{13}{6}h$ 时彻底清洗皮肤, 清空表皮储仓内的 BPS. 在接触实验开始的 $72h$, 受试者被要求每 $4.3h$ 左右提供一次尿液样本, 以检测尿液中的 BPS 与 BPS-g 的含量 (不检测 BPS-s 的尿液含量, 因为缺少相应的检测试剂). 在另一个 BPS 人体接触实验中, Khmiri 等人^[13]同时监测了受试者的血液与尿液. 接触 BPS 起的前 $2h$ 内每 $0.25h$ 取样一次血液, 第 2 小时至第 8 小时内每 $1h$ 取样一次血液, 之后分别在 $10h$, $24h$, $48h$ 时取样一次血液. 尿液的取样节点不是固定的, 而是将接触 BPS 后 $48h$ 分成了 11 个时段. 受试者在单个时段内的所有排尿都会被收集, 作为该时段标签下的一个整体取样.

根据 Hu 等人的文章^[1], 从受试者与热敏纸接触时起, 热敏纸内的 BPS 通过手指表皮储存进入毛囊和角质层, 接着扩散进入活性表皮层, 再通过毛囊和活性表皮层与内环境的交换进入体循环, 平行分层皮肤仓室内的物质交换情况的图片形式如图 2.1. 血浆携带 BPS 通过血液交换将其送入肝脏, 脑, 脂肪, 性腺等组织器官. 部分 BPS 在肝脏反应为 BPS-g 或 BPS-s. 体内的 BPS-g 与 BPS-s 不会再反应为其他物质, 这两种物质会像始终未发生反应的 BPS 一样, 最终随血液进入肾脏, 通过尿液排出. 在只有皮肤暴露途径吸收外源 BPS 的情况下, 不考虑 BPS 的肝肠循环过程, 认为胃肠道不存在 BPS, 且小肠不发生 BPS 葡萄糖酸化反应为 BPS-g 的反应. 整个 PBPK 模型的仓室间 BPS, BPS-g, BPS-s 的交换情况如图 2.2 所示.

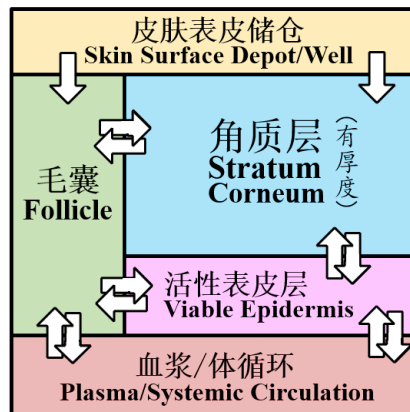


图 2.1: 平行分层皮肤仓室的简化图解 (箭头代表 BPS 的可能运输方向)

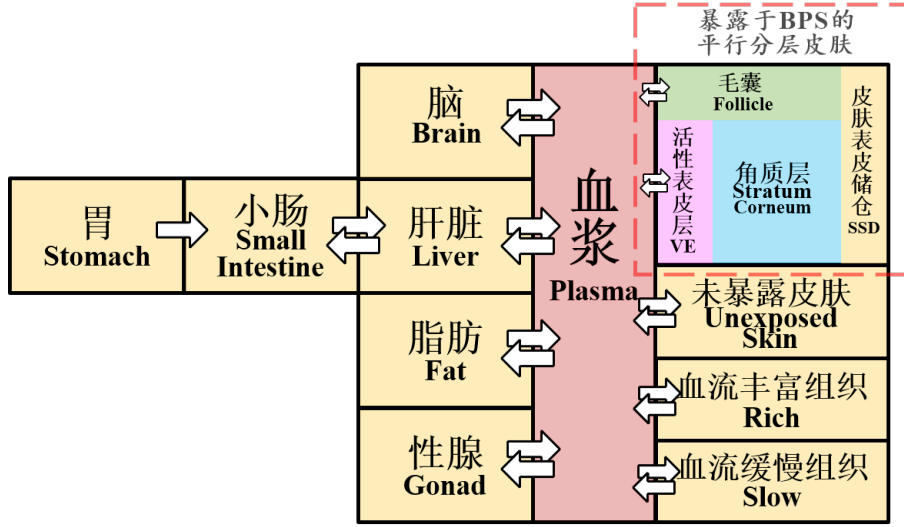


图 2.2: BPS 人体内 PBPK 模型的简化图解 (箭头代表 BPS 的可能运输方向)

2.2 模型的数学形式

本文中的 PBPK 模型共有 14 个仓室，包括：暴露皮肤 {皮肤表皮储仓，角质层，毛囊，活性表皮层}，胃，小肠，未暴露于化学品的皮肤，血浆，脂肪，性腺，肝脏，脑，血流丰富组织，血流缓慢组织。其中暴露的含义为“在实验中与 BPS 接触”。根据各仓室之间的关系以及 BPS 在各仓室内的生化反应，得到 19 个关于时间 t 变化的常微分方程与 1 个偏微分方程。这些微分方程共同构成了 BPS 的带平行分层皮肤仓室的 PBPK 模型。

角质层内 BPS 的浓度 $\varphi(x, t)$

如 (2.1)，其中的偏微分方程本质上是一个扩散对流方程，其解 φ 代表角质层 (Stratum Corneum) 中 BPs 含量，自变量 x 代表角质层的深度，自变量 t 代表时间， DSC 代表 BPS 在角质层中的有效扩散系数 (cm^2/h)， u_1 代表 BPS 随脱屑向皮肤表面转移的速度 (cm/min)， T_{SC} 代表角质层的深度 (um)， HSC_{well} 代表角质层和皮肤表皮储仓之间的分配系数， $C_{well}(t)$ 代表皮肤表皮储仓在 t 时刻的 BPS 浓度 ($nmol/cm^3$)， HSC_{VE} 代表角质层和活性表皮之间的分配系数， $C_{VE}(t)$ 代表活性表皮在 t 时刻的 BPS 浓度 ($nmol/cm^2$)。

$$\begin{cases} \frac{\partial \varphi(x, t)}{\partial t^2} = DSC \frac{\partial^2 \varphi(x, t)}{\partial x^2} + u_1 \frac{\partial \varphi(x, t)}{\partial x}, & 0 \leq x \leq T_{SC}, \quad t \geq 0 \\ \varphi(0, t) = HSC_{well} \times C_{well}(t), & t \geq 0 \\ \varphi(T_{SC}, t) = HSC_{VE} \times C_{VE}(t), & t \geq 0 \\ \varphi(x, 0) = 0, & 0 \leq x \leq T_{SC} \end{cases} \quad (2.1)$$

使用空间离散化的办法，将角质层视作一个长度为 T_{SC} 的线段，将该线段等距分为 10 段，共 11 个节点 $\{x_i\}_{0 \leq i \leq 10}$ 。其中第一个节点 x_0 视作皮肤表皮储仓，最后一个节点 x_{10} 视作活性表皮。 x_1 至 x_9

九个节点处 BPS 浓度的一阶与二阶空间导数值通过中心差分法近似表示:

$$\begin{cases} \frac{\partial^2}{\partial x^2} \varphi(x_j, t) \approx \frac{\varphi_{j+1}(t) - 2\varphi_j(t) + \varphi_{j-1}(t)}{(\Delta x)^2} \\ \frac{\partial}{\partial x} \varphi(x_j, t) \approx \frac{\varphi_{j+1}(t) - \varphi_{j-1}(t)}{2\Delta x} \end{cases} \quad (2.2)$$

整理后得到这九个节点处 BPS 浓度关于时间的一阶导数值如 (2.3), $i = 2, 3, \dots, 8$ 与 (2.4), (2.5). 其中 $SCDX = \frac{T_{SC}}{10}$, V_{well} 为暴露皮肤表面储仓沉积体积 (cm^3), $A_{well}(t)$ 为皮肤表皮储仓在 t 时刻的 BPS 含量 ($nmol$), V_{TVE} 为暴露皮肤活性表皮层体积 (cm^3), $A_{VE}(t)$ 为活性表皮在 t 时刻的 BPS 含量 ($nmol$). 节点 x_1 和 x_9 处的时间导数值与其他节点不同的原因是: 与它们相邻的部分涉及到皮肤的不同层室, 需要利用 BPS 在不同层室组织间的分配系数来确定两个不同层室间 BPS 的转移情况.

$$\begin{aligned} \frac{dC_{SCi}(t)}{dt} = & \left(\frac{DSC}{SCDX^2} - \frac{u_1}{2 \times SCDX} \right) C_{SCi-1}(t) - \frac{2 \times DSC}{SCDX^2} C_{SCi}(t) \\ & + \left(\frac{DSC}{SCDX^2} + \frac{u_1}{2 \times SCDX} \right) C_{SCi+1}(t). \end{aligned} \quad (2.3)$$

$$\begin{aligned} \frac{dC_{SC1}(t)}{dt} = & \left(\frac{DSC \times HSC_{well}}{V_{well} \times SCDX^2} - \frac{u_1 \times HSC_{well}}{V_{well} \times 2 \times SCDX} \right) A_{well}(t) - \frac{2 \times DSC}{SCDX^2} C_{SC1}(t) \\ & + \left(\frac{DSC}{SCDX^2} + \frac{u_1}{2 \times SCDX} \right) C_{SC2}(t). \end{aligned} \quad (2.4)$$

$$\begin{aligned} \frac{dC_{SC9}(t)}{dt} = & \left(\frac{DSC}{SCDX^2} - \frac{u_1}{2 \times SCDX} \right) C_{SC8}(t) - \frac{2 \times DSC}{SCDX^2} C_{SC9}(t) \\ & + \left(\frac{DSC \times HSC_{VE}}{V_{TVE} \times SCDX^2} - \frac{u_1 \times HSC_{VE}}{V_{TVE} \times 2 \times SCDX} \right) A_{VE}(t). \end{aligned} \quad (2.5)$$

接下来介绍模型中余下的 19 个解关于时间 t 变化的常微分方程, 此部分内方程中出现的变量与常量的含义详情见附录 A.

毛囊内 BPS 的含量 $A_{Fo}(t)$

见 (2.6), 方程等式两端为 $A_{Fo}(t)$ 的一阶导数. 在该 PBPK 模型中, 毛囊与皮肤表皮储仓和血液之间有着直接的物质交换, 故毛囊内 BPS 含量的增减与表皮储仓或血浆中 BPS 的含量有关, BPS 在宏观上遵循着顺浓度梯度运输的规则在不同组织内交换. 同时, 毛囊内 BPS 含量的增减会受到自身的限制, 当毛囊内 BPS 浓度高于血浆或表皮储仓中 BPS 浓度时, BPS 会顺浓度梯度进入浓度小的组织.

$$\begin{aligned} \frac{dA_{Fo}(t)}{dt} = & - \left(\frac{Pfo \times AEXP \times FEXP}{V_{TFo} \times HFo_{well}} + \frac{Qskin \times AEXP \times 0.25}{BSA \times V_{TFo} \times pskin} \right) A_{Fo}(t) \\ & + \frac{Pfo \times AEXP \times FEXP}{V_{well}} A_{well}(t) + \frac{Qskin \times AEXP \times 0.25}{BSA \times V_{plasma}} A_{plasma}(t). \end{aligned} \quad (2.6)$$

皮肤表皮储仓内 BPS 的含量 $A_{well}(t)$

见 (2.7), 方程等式两端为 $A_{well}(t)$ 的一阶导数. 类似 (2.6), 等式右端表现出了表皮储仓与节点 x_1 处的角质层和毛囊之间的顺浓度梯度物质交换关系. 同时, 皮肤表皮储存作为在实验中

直接和外源 BPS 接触的部位, 它要接受一个剂量为 $f_1(t)$ (单位:nmol) 的持续的外源 BPS 输入. 当 $t > Time_{add} = \frac{1}{6}h$ 时, 皮肤停止接触外源 BPS, $f_1(t) = 0$. 等式右端有一个因数 $ON(t)$, 当 $t \leq Time_{expose} = \frac{13}{6}h$ 时, $ON(t) = 1$, 皮肤处于 BPS 暴露状态; 当 $t > Time_{expose}$ 时, $ON(t) = 0$, 暴露过 BPS 的皮肤被彻底清洗, 表皮储仓清空, 后续 $A_{well}(t)$ 的值与一阶导数值都为 0.

$$\begin{aligned} \frac{dA_{well}(t)}{dt} = & \left(\frac{DSC \times AEXP \times (1 - FEXP)}{SCDX} C_{SC1}(t) - \frac{Pfo \times AEXP \times FEXP}{V_{TFO} \times HFO_{well}} A_{Fo}(t) \right. \\ & - \left(\left(\frac{DSC \times HSC_{well}}{V_{well} \times SCDX} - \frac{u_1 \times HSC_{well}}{V_{well}} \right) \times AEXP \times (1 - FEXP) \right. \\ & \left. \left. - \frac{Pfo \times AEXP \times FEXP}{V_{well}} \right) A_{well}(t) + f_1(t) \right) \times ON(t). \quad (2.7) \end{aligned}$$

活性表皮层内 BPS 的含量 $A_{VE}(t)$

见 (2.8), 方程等式两端为 $A_{VE}(t)$ 的一阶导数. 类似 (2.6), 等式右端表现出了活性表皮层与节点 x_9 处的角质层, 毛囊和血浆之间的顺浓度梯度物质交换关系.

$$\begin{aligned} \frac{dA_{VE}(t)}{dt} = & \frac{DSC \times AEXP \times (1 - FEXP)}{SCDX} C_{SC9}(t) + \left(\left(\frac{-DSC \times HSC_{VE}}{V_{TVE} \times SCDX} - \frac{u_1 \times HSC_{VE}}{V_{TVE}} \right) \times \right. \\ & \left. AEXP \times (1 - FEXP) - \frac{Qskin \times AEXP \times 0.75}{BSA \times V_{TVE} \times pskin} \right) A_{VE}(t) + \frac{Qskin \times AEXP \times 0.75}{BSA \times V_{plasma}} A_{plasma}(t). \quad (2.8) \end{aligned}$$

胃部 BPS 的含量 $A_{ST}(t)$

见 (2.9), 方程等式两端为 $A_{ST}(t)$ 的一阶导数. 等式右端的第一个加数代表胃部 BPS 向肝脏和小肠转移的过程, 第二个加数是剂量为 $f_2(t)$ (单位:nmol) 的通过口服吸收的外源 BPS. 但由于本文不考虑口服吸收 BPS 的情形, $f_2(t) \equiv 0$, 可认为胃部内始终不含有 BPS 或它的衍生物.

$$\frac{dA_{ST}(t)}{dt} = -(k_0 + ge)A_{ST}(t) + f_2(t). \quad (2.9)$$

未暴露于 BPS 的皮肤, 脂肪, 性腺, 脑部, 血流丰富组织, 血流缓慢组织的 BPS 的含量

见 (2.10) 至 (2.15), 各方程的等式两端为小标题中组织器官内 BPS 含量的一阶导数. 这些等式的右端一致地代表了对应组织器官与血浆之间的顺浓度梯度物质交换关系.

$$\frac{dA_{skin}(t)}{dt} = \frac{-Qskin \times (1 - \frac{AEXP}{BSA})}{(V_{skin} - V_{TSC} - V_{TVE} - V_{TFO}) \times pskin} A_{skin}(t) + \frac{Qskin \times (1 - \frac{AEXP}{BSA})}{V_{plasma}} A_{plasma}(t). \quad (2.10)$$

$$\frac{dA_{fat}(t)}{dt} = \frac{-Qfat}{V_{fat} \times pfat} A_{fat}(t) + \frac{Qfat}{V_{plasma}} A_{plasma}(t). \quad (2.11)$$

$$\frac{dA_{gonad}(t)}{dt} = \frac{-Qgonad}{V_{gonad} \times pgonad} A_{gonad}(t) + \frac{Qgonad}{V_{plasma}} A_{plasma}(t). \quad (2.12)$$

$$\frac{dA_{brain}(t)}{dt} = \frac{-Qbrain}{V_{brain} \times pbrain} A_{brain}(t) + \frac{Qbrain}{V_{plasma}} A_{plasma}(t). \quad (2.13)$$

$$\frac{dA_{rich}(t)}{dt} = \frac{-Q_{rich}}{V_{rich} \times prich} A_{rich}(t) + \frac{Q_{rich}}{V_{plasma}} A_{plasma}(t). \quad (2.14)$$

$$\frac{dA_{slow}(t)}{dt} = \frac{-Q_{slow}}{V_{slow} \times pslow} A_{slow}(t) + \frac{Q_{slow}}{V_{plasma}} A_{plasma}(t). \quad (2.15)$$

血浆内的 BPS 的含量

见 (2.16), 方程等式两端为血浆内 BPS 含量的一阶导数. 血浆是体循环的重要组成部分, 该 PBPK 模型内的各仓室由血浆连接起来, 几乎每个仓室都与血浆有直接的物质交换, BPS 从组织器官内进入血浆, 血浆又携带着 BPS 进入各个组织器官. 等式的右端代表了血浆与毛囊, 活性表皮层, 未直接暴露于 BPS 的皮肤组织, 脂肪, 性腺, 脑部, 血流丰富组织, 血流缓慢组织, 肝脏之间的顺浓度梯度物质交换关系. 等式右端除了 $A_{plasma}(t)$ 项外每一项都有因数 $Qc - Kurinebps$, 其中 Qc 是心脏血流速度 (L/h), $Kurinebps$ 是 BPS 的尿液排泄参数 (L/h), 它们相减代表了每轮血液循环净剩的携带了 BPS 的血浆量.

$$\begin{aligned} \frac{dA_{plasma}(t)}{dt} = & \frac{(Qc - Kurinebps) \times Q_{skin} \times \frac{AEXP}{BSA} \times 0.25}{Qc \times V_{TFo} \times pskin} A_{Fo}(t) \\ & + \frac{(Qc - Kurinebps) \times Q_{skin} \times \frac{AEXP}{BSA} \times 0.75}{Qc \times V_{TVE} \times pskin} A_{VE}(t) \\ & + \frac{(Qc - Kurinebps) \times Q_{skin} \times (1 - \frac{AEXP}{BSA})}{Qc \times (V_{skin} - V_{TSC} - V_{TVE} - V_{TFo}) \times pskin} A_{skin}(t) \\ & + \frac{(Qc - Kurinebps) \times Q_{fat}}{Qc \times V_{fat} \times pfat} A_{fat}(t) + \frac{(Qc - Kurinebps) \times Q_{gonad}}{Qc \times V_{gonad} \times pgonad} A_{gonad}(t) \\ & - \frac{Qc}{V_{plasma}} A_{plasma}(t) + \frac{(Qc - Kurinebps) \times Q_{brain}}{Qc \times V_{brain} \times pbrain} A_{brain}(t) \\ & + \frac{(Qc - Kurinebps) \times Q_{rich}}{Qc \times V_{rich} \times prich} A_{rich}(t) + \frac{(Qc - Kurinebps) \times Q_{slow}}{Qc \times V_{slow} \times pslow} A_{slow}(t) \\ & + \frac{(Qc - Kurinebps) \times Q_{liver}}{Qc \times V_{liver} \times pliver} A_{liver}(t). \end{aligned} \quad (2.16)$$

胃肠部的 BPS-g 的含量 $A_{GIBPSg}(t)$ 与胃肠部的 BPS-s 的含量 $A_{GIBPSs}(t)$

见 (2.17), 方程等式两端为胃肠部的 BPS-g 的含量的一阶导数. 等式右端的第二个加数代表了小肠内的 BPS 葡萄糖酸化为 BPS-g 的过程, 第一个加数代表了 BPS-g 从胃肠部进入血液的过程. 在本文中, 没有口服吸收途径, 认为胃肠部不存在 BPS-g 或 BPS-s. 对于 (2.18) 有类似的说明, 不同在于小肠内几乎不发生 BPS 硫酸盐化为 BPS-s 的反应.

$$\frac{dA_{GIBPSg}(t)}{dt} = -kGIing \times A_{GIBPSg}(t) + \frac{Vmaxgutg \times A_{SI}(t)}{enterocytes \times Kmgutg + A_{SI}(t) + \frac{A_{SI}(t)^2}{enterocytes \times Ksigutg}}. \quad (2.17)$$

$$\frac{dA_{GIBPSs}(t)}{dt} = -kGIins \times A_{GIBPSs}(t) + \frac{Vmaxguts \times A_{SI}(t)}{enterocytes \times Kmguts + A_{SI}(t)}. \quad (2.18)$$

小肠的 BPS 的含量 $A_{SI}(t)$

见 (2.19), 方程等式两端为小肠的 BPS 的含量的一阶导数. 等式右端的第一个加数代表了口服 BPS 时胃部 BPS 进入小肠的过程, 第二个加数代表了小肠内部分 BPS 进入肝脏的过程, 后两个加数代表了部分 BPS 在小肠内葡萄苷酸化为 BPS-g 与硫酸盐化为 BPS-s 的反应过程 (小肠内几乎不发生 BPS 磷酸盐化反应, 此处对应的最大反应速率 $Vmaxguts$ 非常小). 在本文中, 没有口服吸收 BPS 的途径, 认为小肠内不存在 BPS 及其衍生物.

$$\begin{aligned} \frac{dA_{SI}(t)}{dt} = & ge \times A_{ST}(t) - k1 \times A_{SI}(t) - \frac{Vmaxgutg \times A_{SI}(t)}{enterocytes \times Kmgutg + A_{SI}(t) + \frac{A_{SI}(t)^2}{enterocytes \times Ksigutg}} \\ & - \frac{Vmaxguts \times A_{SI}(t)}{enterocytes \times Kmguts + A_{SI}(t)}. \quad (2.19) \end{aligned}$$

肝脏的 BPS 的含量 $A_{liver}(t)$

见 (2.20), 方程等式两端为肝脏的 BPS 的含量的一阶导数. 等式右端的前两个加数分别代表了口服 BPS 时胃部和小肠内 BPS 进入肝脏的过程, 第三第四个加数代表了肝脏与血浆之间的顺浓度梯度物质交换关系, 第五第六个加数代表了 BPS-g 与 BPS-s 引起肝脏内部分 BPS 发生肝肠循环的过程, 后两个加数代表了部分 BPS 在肝脏内葡萄苷酸化为 BPS-g 与硫酸盐化为 BPS-s 的反应过程.

$$\begin{aligned} \frac{dA_{liver}(t)}{dt} = & k0 \times A_{ST}(t) + k1 \times A_{SI}(t) + \frac{Qliver}{V_{plasma}} A_{plasma}(t) - \frac{Qliver}{V_{liver} \times pliver} A_{liver}(t) \\ & + kenterobpsg \times A_{BPSg_delay}(t) + kenterobpss \times A_{BPSs_delay}(t) \\ & - \frac{Vmaxliverg \times A_{liver}(t)}{V_{liver} \times pliver \times Kmliverg + A_{liver}(t)} - \frac{Vmaxlivers \times A_{liver}(t)}{V_{liver} \times pliver \times Kmlivers + A_{liver}(t)}. \quad (2.20) \end{aligned}$$

发生肝肠循环/小肠内的 BPS-g/BPS-s 的量 $A_{BPSg_delay}(t)/A_{BPSs_delay}(t)$

见 (2.21), 方程等式两端为发生肝肠循环的 BPS-g 的量的一阶导数. 等式右端的第一个加数代表了胃肠道 BPS-g 进入血浆后再进入肝肠循环的过程, 第二个加数代表了肝肠循环中的 BPS-g 进入血液循环的过程, 最后一个加数代表了肝脏中 BPS 新转化成的 BPS-g 进入肝肠循环的过程. 对于发生肝肠循环的 BPS-s 的量的方程 (2.22), 有着完全一致的描述.

$$\begin{aligned} \frac{dA_{BPSg_delay}(t)}{dt} = & met2g \times kGIing \times A_{GIBPSg}(t) \\ & - (kentero + k4_{IV} + kenterobpsg) \times A_{BPSg_delay}(t) + \frac{met2g \times Vmaxliverg \times A_{liver}(t)}{V_{liver} \times pliver \times Kmliverg + A_{liver}(t)}. \quad (2.21) \end{aligned}$$

$$\begin{aligned} \frac{dA_{BPSs_delay}(t)}{dt} = & met2s \times kGIins \times A_{GIBPSs}(t) \\ & - (kentero + k4_{IV} + kenterobpss) \times A_{BPSs_delay}(t) + \frac{met2s \times Vmaxlivers \times A_{liver}(t)}{V_{liver} \times pliver \times Kmlivers + A_{liver}(t)}. \quad (2.22) \end{aligned}$$

人体内 BPS-g/BPS-s 的总含量 $A_{BPSg}(t)/A_{BPSs}(t)$

见 (2.23), 方程等式两端为人体内 BPS-g 的总含量的一阶导数. 等式右端的第一个加数代表了口服摄入 BPS 时, 胃肠部产生的 BPS-g 进入血液后未进入肝肠循环的部分 BPS-g, 第二个加数代表了正在进行肝肠循环的 BPS-g, 第三个加数代表了 BPS-g 随尿液排出人体的过程, 最后一个加数代表了 BPS 通过肝脏转化成的 BPS-g 进入血液后未进入肝肠循环的部分 BPS-g. (2.24) 与之相似.

$$\begin{aligned} \frac{dA_{BPSg}(t)}{dt} = & met1g \times kGIing \times A_{GIBPSg}(t) + kentero \times A_{BPSg_delay}(t) - \frac{K_{urinebpsg}}{V_{bodyg} + 10^{-34}} A_{BPSg}(t) \\ & + \frac{met1g \times V_{maxliverg} \times A_{liver}(t)}{V_{liver} \times pliver \times K_{mliverg} + A_{liver}(t)}. \end{aligned} \quad (2.23)$$

$$\begin{aligned} \frac{dA_{BPSs}(t)}{dt} = & met1s \times kGIins \times A_{GIBPSs}(t) + kentero \times A_{BPSs_delay}(t) - \frac{K_{urinebps}}{V_{bodyg} + 10^{-34}} A_{BPSs}(t) \\ & + \frac{met1s \times V_{maxlivers} \times A_{liver}(t)}{V_{liver} \times pliver \times K_{mlivers} + A_{liver}(t)}. \end{aligned} \quad (2.24)$$

综上, 本文的 PBPK 模型的数学形式的主要部分为一个由 28 个常微分方程构成的微分方程组, 其中每个方程的解关于时间 t 变化, 且每个解的初值都是 0. 除此之外, 根据实验过程, 还需计算尿液中的 BPS 与 BPS-g 的在 t 时刻的含量 $A_{urinebps}(t)$ 与 $A_{urinebpsg}(t)$, (2.25) 与 (2.26) 给出了它们的一阶导数值. 这两个等式都不是常微分方程.

$$\begin{aligned} \frac{dA_{urinebps}(t)}{dt} = & K_{urinebps} \times \left(\frac{Q_{skin} \times \frac{AEXP}{BSA} \times 0.25}{Q_c \times V_{Tfo} \times p_{skin}} A_{Fo}(t) + \frac{Q_{skin} \times \frac{AEXP}{BSA} \times 0.75}{Q_c \times V_{TVE} \times p_{skin}} A_{VE}(t) \right. \\ & + \frac{Q_{skin} \times (1 - \frac{AEXP}{BSA})}{Q_c \times (V_{skin} - V_{TSC} - V_{TVE} - V_{Tfo}) \times p_{skin}} A_{skin}(t) \\ & + \frac{Q_{fat}}{Q_c \times V_{fat} \times p_{fat}} A_{fat}(t) + \frac{Q_{gonad}}{Q_c \times V_{gonad} \times p_{gonad}} A_{gonad}(t) \\ & + \frac{Q_{brain}}{Q_c \times V_{brain} \times p_{brain}} A_{brain}(t) + \frac{Q_{rich}}{Q_c \times V_{rich} \times p_{rich}} A_{rich}(t) \\ & \left. + \frac{Q_{slow}}{Q_c \times V_{slow} \times p_{slow}} A_{slow}(t) + \frac{Q_{liver}}{Q_c \times V_{liver} \times p_{liver}} A_{liver}(t) \right). \end{aligned} \quad (2.25)$$

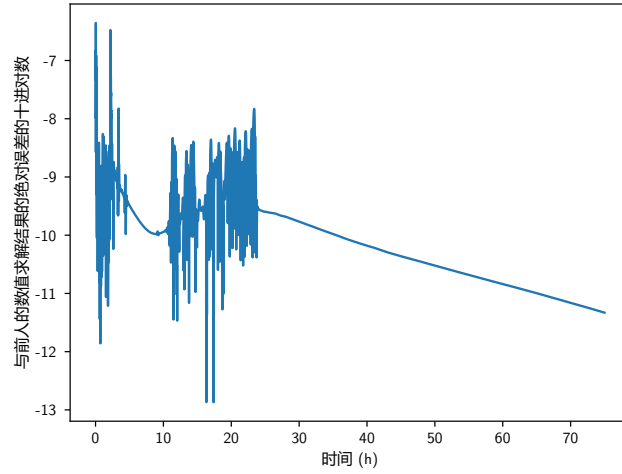
$$\frac{dA_{urinebpsg}(t)}{dt} = \frac{K_{urinebpsg}}{V_{bodyg} + 10^{-34}} A_{BPSg}(t). \quad (2.26)$$

2.3 PBPK 模型的求解

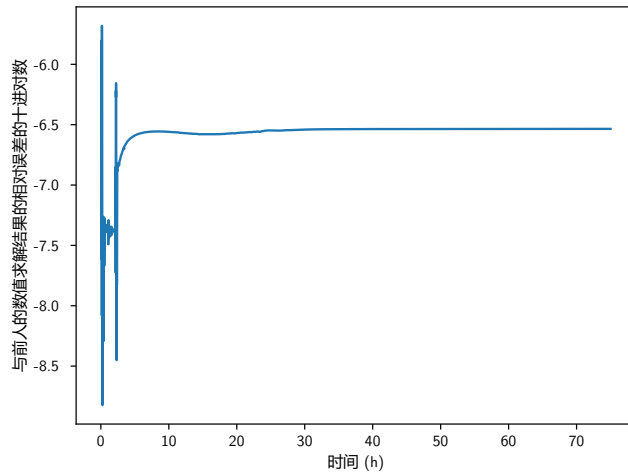
确定模型的数学形式后, 将 Hu 等人^[12]提供的数据相对应地代入至方程的各个参数, 使用 Python 中的第三方库函数 `scipy.integrate.odeint()` 对微分方程组求解, 该函数默认使用 LSODA 数值格式. LSODA 是一种高效的支持自适应步长的算法^[14], 根据设置好的误差限度, 若按照原有的预设步长来进行下一节点的数值计算, 得到的误差大于限度, 则将此步预设步长拆分为更小的多个步长, 以此来减小误差至限度以内; 若按照预设步长计算时得到的误差远小于限度, 则适当增大步

长以减小资源消耗. 同时, LSODA 同时支持刚性与非刚性问题, 处理刚性问题时, 它使用后向微分公式 (BDF), 处理非刚性问题时, 它使用 Runge-Kutta 公式.

根据 Hu 等人^[1]的实验, 每个解的初值都设置为 0, 时间网格设置为 $range(0, 75, 0.005)$, 时间步长为 $0.005h$, 共 15000 个时间节点. 每个方程的解都对应了一条人体某组织器官内 BPS(或 BPS-g, BPS-s) 的含量 ($mmol$) 或浓度 ($mmol/L$) 随时间变化的曲线.



(a) $A_{plasma}(t)$ 的相对误差的十进对数



(b) $A_{urinebps}(t)$ 的相对误差的十进对数

图 2.3: 与前人 PBPK 模型计算出的结果的相对误差

图2.3展示了本文的模型数值求解结果与参考代码^[12]中结果中 $A_{plasma}(t)$ 与 $A_{urinebps}(t)$ 的相对误差, 可以看到两种绝对误差曲线的数值都来到了 10^{-7} 这一量级附近, 说明对 Hu 等人^[1]提出的模型的求解复现效果较好. 分别运行 100 次前人模型求解代码与本文复现的模型求解代码, 取单次运行的平均, 得到本文的模型求解代码的单次平均运行时间为 $0.1086s$, 前人的代码单次平均运行时间为 $18.1621s$, 本文复现代码的运行速度约为前人速度的 167 倍.

第三章 PBPK 模型参数反演的神经网络方法：使用固定时间节点数据

本章使用了对偶学习的办法，使用 PBPK 模型正向生成的不含时间信息的数据分别训练了拟合 PBPK 模型正向求解模型的神经网络以及 PBPK 模型参数反演的神经网络，并展示了神经网络在测试数据上的效果。本章将讲解数据集的结构与构造以及网络模型的结构，训练与效果。

3.1 参数反演的背景与网络模型的基本介绍

Hu 等人^[1]在原本人体关于 BPs 的 PBPK 模型的基础上创新了平行分层仓室模型，该部分仓室模型在参数的数值确定方面还不完善，需要借助人体实验得到的真实数据对该部分参数进行反演校准。需要反演校准的与皮肤仓室有关的参数有三个，称由这三个参数的取值组成的一个数组为皮肤三参数组合，它们分别为：

$$\left| \begin{array}{ll} DSC & : \text{角质层中的有效扩散系数 } (cm^2/h) \\ u_1 & : \text{由脱屑而向皮肤表面转移的速度 } (cm/min) \\ Pfo & : \text{毛囊的渗透系数 } (cm/h) \end{array} \right|$$

如实验背景介绍中的内容，受试者在接受 BPS 的皮肤接触后，需要在实验进行的 75h 内收集每次排出的尿液^[1]。研究者需测量尿液中 BPS 与 BPS-g 的累计含量 ($mmol$)，具体做法为：对于单次排尿，首先记录排尿时间，然后测量本次排出尿液的化学品含量，对截止至本次（包括本次）的全部排尿中的化学品含量累计求和，将得到的求和数值作为本次排尿时间节点处的受试者尿液中 BPS 或 BPS-g 的累计含量，这样就得到了单次排尿的（排尿时间 t_i ，尿液累计化学品含量 A_i ）。将这样得到的所有数据点绘制在含量-时间坐标系内，可得到一幅散点图。

为了校准模型参数，需要使 PBPK 模型计算出的尿液中 BPS 与 BPS-g 含量之和的曲线拟合散点图中的散点。Hu 等人^[1]设置损失函数为真实受试者尿液化学品含量与 PBPK 模型结果中的尿液化学品含量之间的 MSE ，以三个待优化的参数为目标，利用 R 语言中的优化函数 `optim()` 最小化损失函数。函数的输出即为参数的校准值。

另外，根据 Khmiri 等人^[13]的研究，对受试者在固定时间节点采集血浆样本，并分析其中的 BPS 含量，将损失函数设置为真实受试者血药含量与 PBPK 模型结果中的血药含量之间的 MSE ，也是一种参数校准的手段。相对于人体其他组分，血浆或尿液更容易获取与分析化学性质。两种手段都以 PBPK 模型的输出类型之一为参照物，通过实验获得真实数据，与模型输出数据进行对比，以此来指导模型内部分参数的校准。

给定血药含量或尿药含量关于时间变化曲线上的散点，反演得到模型内的三个参数。基于这样的参数反演流程，本文使用了神经网络模型来辅助 PBPK 模型的参数校准，在传统优化算法之外提供了一种更有效且更快速的方法。本章介绍的是基于 Hu 等人^[1]提出的 PBPK 模型与模型参数数值，使用固定时间采样点数据训练的皮肤三参数反演神经网络模型。其中固定时间采样点的意思

是, 假设在理想情况下, 研究者能严格按照某组合理的预设时间节点采集受试者的血液与尿液. 该神经网络模型可以接收在固定时间点处采集的血浆和尿液中的 BPS 或 BPS-g 的含量, 也可以接收 PBPK 模型输出的血药含量, 尿药含量曲线在固定时间点处的采样点, 网络模型输出反演后的皮肤三参数组合. 该网络的数据集结构并未完全参考 Hu 等人^[1]的参数校准过程, 利用真实数据校准参数的网络模型将会在下一章介绍.

3.2 参数反演神经网络模型的数据集结构与构建

3.2.1 网络功能与数据集结构介绍

数据集的结构直接决定了网络的输入与输出数据的形状, 间接决定了网络的功能与效果, 同时网络功能的实现又需要设计合适的数据集结构, 故在此一齐介绍.

首先, 认为 PBPK 模型中除了皮肤三参数组合外的其他参数都已确定为常数, 并固定求解模型中微分方程组的时间格点与初值, 设置皮肤三参数组合为 PBPK 模型的求解模型的唯一输入, 将整个 PBPK 模型的求解模型抽象为一个函数, 如 (3.1) 所示, $\vec{\alpha}$ 代表皮肤三参数组合, \vec{A}_{BPS} 代表某一组分中的 BPS 含量在给定固定时间节点上的数值, 相应地, n 是时间节点组的节点数量,

$$f: (\mathbb{R}^3)^+ \rightarrow (\mathbb{R}^n)^+, \quad \vec{\alpha} \mapsto \vec{A}_{BPS}. \quad (3.1)$$

本章内介绍的参数反演神经网络的功能就是拟合 (3.1) 的反函数, 输入固定节点处血浆与尿液内 BPS 或 BPS-g 的含量数值构成的向量 $(A(t_1), A(t_2), \dots, A(t_n))$, 网络模型会输出一组近似的与输入对应的皮肤三参数组合 $\vec{\alpha}^*$.

数据集由特征 (feature) 与标签 (label) 构成, 特征是神经网络模型的输入, 标签是网络模型输出的对照物, 网络模型的输出是对标签的拟合. 一条数据包含一条特征和一条标签, 二者之间是绑定关系. 在本章的参数反演神经网络模型的数据集中, 一条特征是一个含有 n 个元素的一维数组, 实际含义是 PBPK 模型输出的 $A_{plasma}(t)$, $A_{urinebps}(t)$, $A_{urinebpsg}(t)$ 在固定时间节点 $\{t_i^{plasma}\}_{1 \leq i \leq n_1}$, $\{t_i^{urine}\}_{1 \leq i \leq n_2}$ 上的取样值. 一条标签是一个含有 3 个元素的一维数组, 实际含义是与代表 PBPK 模型输出的特征对应的皮肤三参数组合 $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$. 整个数据集由大量数据组成, 设置数据条数为 N , 则整个特征集是一个形状为 $N \times n$ 的二维数组, 通常记为 X ; 与之对应的标签集是一个形状为 $N \times 3$ 的二维数组, 通常记为 y .

3.2.2 标签集的构建

4.2 标签集, 一个 $N \times 3$ 的二维数组, 由 N 条皮肤三参数组合组成. 由于 DSC , u_1 , Pfo 这三个参数的数量级较小, 为了更方便地表示其数值, 将标签 $\vec{\alpha}$ 设置如下: 在原参数数值前乘以一个系数, 使得调整后的参数数量级处在 10^0 或 10^1 ,

$$\begin{aligned} \alpha_1 &= DSC \times 10^9, \\ \alpha_2 &= u_1 \times 10^5, \\ \alpha_3 &= Pfo \times 10^5. \end{aligned}$$

本文使用了两种方式得到大量不同的皮肤三参数组合 $\vec{\alpha}$ 数据, 分别是依概率分布随机取样以及稀疏网格法 (Sparse Grids). 之后将两种方法得到的数据合并在一起, 构成本章神经网络使用的标签集. 以下是两种构建方法的具体介绍.

截断正态分布取样

Hu 等人^[1]提出了该种取样方法, 并提供了皮肤三参数 α_i 的平均值 $mean_i$, 并设定每个参数的变异系数 $CV_i = 0.3$, 通过公式 $std = CV \times mean$ 可得到三个参数的标准差 std_i . 分别建立三个参数的截断正态分布, 分布的平均值为 $mean_i$, 标准值为 std_i , 左截断点为 $lb_i = mean_i - 1.96 \times std_i$, 右截断点为 $ub_i = mean_i + 1.96 \times std_i$. 确定截断正态分布后, 使用 Python 第三方库函数 `scipy.stats.truncnorm()` 根据三种概率分布对三个参数进行独立的采样, 采样 α_1 共 32 个, α_2 共 30 个, α_3 共 30 个. 接下来取它们的笛卡尔积, 得到 $32 \times 30 \times 30 = 28,800$ 组皮肤三参数组合 $\vec{\alpha}$.

稀疏网格法

稀疏网格法^[15-16]是一种可用于在高维参数空间取样的高效方法. 以 m 维参数空间下的同维度等距的网格为例, 有如下的稀疏网格生成迭代步骤:

1. 设置复杂度 $k \in \mathbb{N}^+$, 设置参数取样的平均值或中心点 $\bar{x} = (\bar{x}_1, \dots, \bar{x}_m)$ 以及每个维度取样的上下界 $\{u_i, l_i\}_{1 \leq i \leq m}$ (应满足 $\frac{u_i + l_i}{2} = \bar{x}_i$). 第 i 个维度的网络格点间距 $\Delta x_i = (u_i - l_i) \times 2^{-k}$. 如此, 便得到了一个 m 维空间上含有 $(2^k + 1)^m$ 个格点的网格, 称该网格为密网格.
设置迭代步 p , 初始 $p = 1$. 初始状态下稀疏网格只有中心点 \bar{x} 一个点, 设置 \bar{x} 为起点 x_{start} , 加入至起点集 S .
2. 若 $2^{k-p} - 1 < 1$, 则取样结束. 否则, 向下执行.
3. $\forall x_{start} \in$ 起点集 S , 从该点出发, 向每个维度的正负方向共 $2 \times m$ 个方向上分别以 $2^{(p-1)} \cdot \Delta x_i$ 为步长连续不间断地在密网格上取 $2^{k-p} - 1$ 个格点加入到稀疏网格中. 例如, 在第 i 个维度的正方向上取

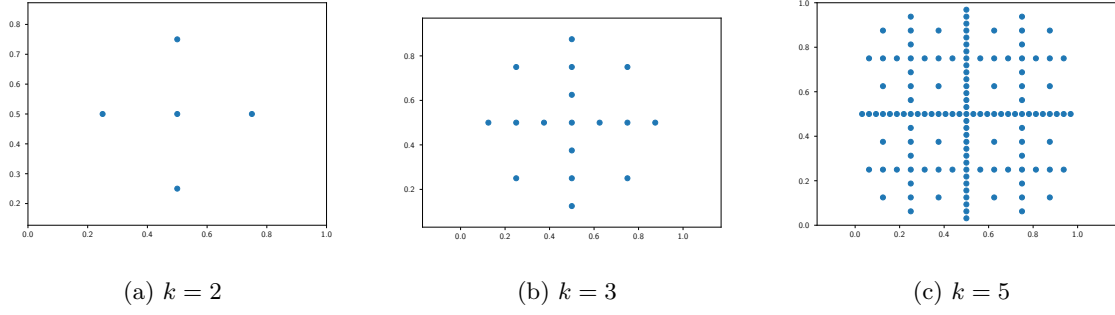
$$\{(x_{start}^1, \dots, x_{start}^{i-1}, x_{start}^i + j \cdot 2^{(p-1)} \cdot \Delta x_i, x_{start}^{i+1}, \dots, x_{start}^m)\}_{1 \leq j \leq 2^{k-p} - 1}.$$

当迭代步数 $p \geq 2$ 时, 若密网格在 x_{start} 的某方向上 (不妨设为第 q 个维度) 有与 x_{start} 距离小于等于 $2^{(p-1)} \cdot \Delta x_q$ 的点已被加入至稀疏网格中, 则不在 x_{start} 的该方向上取点, 并称该方向为 x_{start} 的未取样方向. 在该步骤中, 会有被重复放入稀疏网格中的密网格上的点, 在稀疏网格中视作同一元素.

4. 当 $p = 1$ 时, $\forall 1 \leq i \leq m$, 将中心点 \bar{x} 在第 i 维度正负两个方向上与之距离为 $(u_i - l_i) \times 2^{-(p+1)}$ 的两个点加入至起点集中 S , 共有 $2 \times m \times 2^{m \times (p-1)}$ 个新的起点被加入至起点集 S 中, 删去起点集中除了新起点外的起点.

当 $p \geq 2$ 时, $\forall x_{start} \in$ 起点集 S , 从该点出发, 找到并记录稀疏网格在 x_{start} 除未取样方向的方向上与之距离为 $(u_i - l_i) \times 2^{-p}$ 的点, 加入至分形中心点集 M . 在此过程中, 会有被重复放入 M 的密网格上的点, 在 M 中视作同一元素. 接下来, $\forall x_{middle} \in$ 分形中心点集 M , 从该点出发, 找到并记录稀疏网格在 x_{middle} 在每个维度正负两个方向共 $2 \times m$ 个方向上与之距离为 $(u_i - l_i) \times 2^{-(p+1)}$ 的点, 加入至起点集 S 中. 共有 $2 \times m \times 2^{m \times (p-1)}$ 个新的起点被加入至起点集 S 中, 删去起点集中除了新起点外的起点.

5. $p = p + 1$, 并跳转至步骤 2.

图 3.1: 不同复杂度 k 下的二维稀疏网络

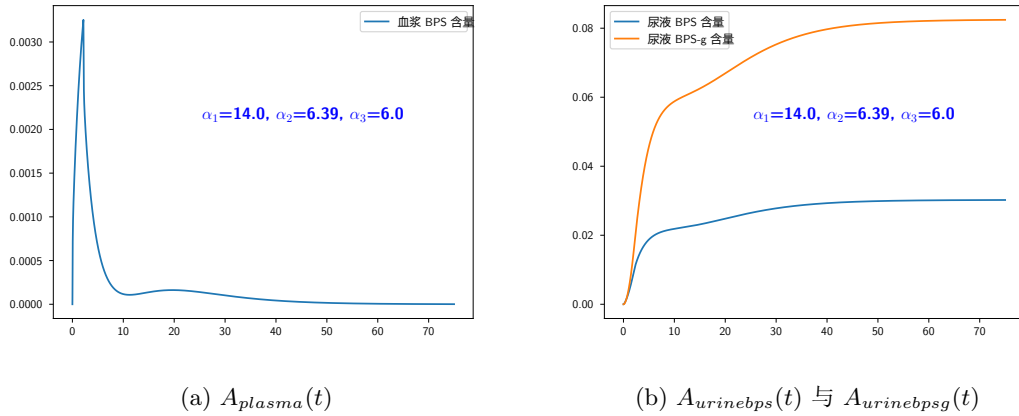
该部分标签集的构造使用了复杂度 $k = 9$ 的三维稀疏网格法, 其中中心点为 $(mean1, mean2, mean3)$, $u_1 - l_1 = 14$, $u_2 - l_2 = 10$, $u_3 - l_3 = 9$, 这些数值的设定参考了截断正态分布取样法. 通过稀疏网格法共得到了 18943 条皮肤三参数组数据. 两种方法得到的标签集拼接起来共 47743 条数据, 整个标签集的形状为 47743×3 .

3.2.3 特征集的构建

?? 一条特征和一条标签互相唯一对应, 本章中的参数反演神经网络模型使用的特征集也有 47743 条. 对于标签集矩阵的每一行的皮肤三参数组, 将其代入至 PBPK 模型的求解模型, 得到血浆中 BPS 的含量 $A_{plasma}(t)$, 尿液中 BPS 和 BPS-g 的累计含量 $A_{urinebps}(t)$, $A_{urinebpsg}(t)$ 关于时间变化的曲线. 为了适配人体实验中的真实采样数据, 模拟真实采样情景, 需要对曲线做采样. 以下介绍了两种特征集的构造, 将分别用于训练两种不同的参数反演神经网络模型.

血药含量采样 28 点 & 尿药含量采样 15 点

此小节的采样时间点设置参考了前人人体实验的设定^[1,13], 且结合了曲线的特性. 如图3.2所示, 曲线的总时间轴是 $0 \sim 75h$, 相较于其他时间段, $A_{plasma}(t)$ 在 $0 \sim 10h$ 的变化十分强烈, 需要在该时间段内取样额外多的时间点, 以更好地捕捉到曲线的特征.

图 3.2: $\vec{\alpha} = (14, 6.39, 6)$ 时 PBPK 模型输出的部分曲线图像

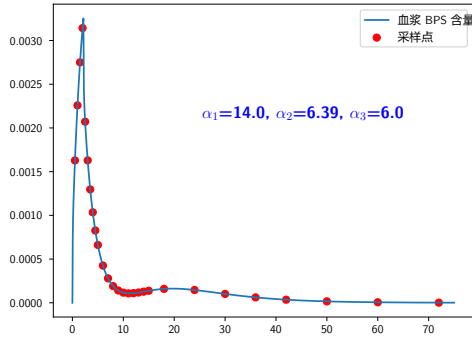
一条血药含量曲线上共采样 28 个时间点 (单位: h):

(0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 24, 30, 36, 42, 50, 60, 72).

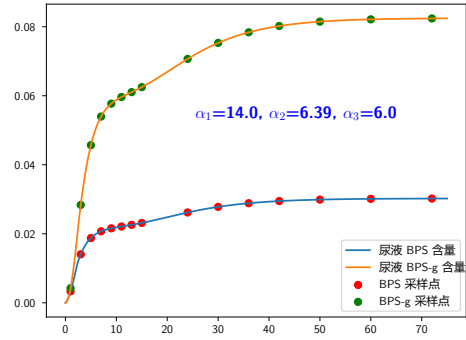
两条尿药含量曲线上分别采样相同的 15 个时间点 (单位: h):

(1, 3, 5, 7, 9, 11, 13, 15, 24, 30, 36, 42, 50, 60, 72).

图3.3展示了三条含量曲线的采样点的位置. 一条标签对应的三条曲线在固定时间节点处共采样了 58 个化学品含量数据, 一条特征即为含有 58 个元素的一维数组, 对应的标签为 PBPK 模型的输入皮肤三参数组, 整个特征集为形状是 47743×58 的二维数组.



(a) $A_{plasma}(t)$ 与 28 点采样点



(b) $A_{urinebps}(t), A_{urinebpsg}(t)$ 与 15 点采样点

图 3.3: $\vec{\alpha} = (14, 6.39, 6)$ 时 PBPK 模型输出的部分曲线与采样点的图像 A

血药含量采样 5 点 & 尿药含量采样 5 点

该部分特征集的构建欲探究减少采样点的可行性. 如图3.2a所示, $A_{plasma}(t)$ 的图像有两个极大值和一个极小值, 第一个极大值也是最大值 MAX_{plasma} . 最大值点出现在曲线变化幅度大的时段, 三个极值点都处在非平缓的时间段. 对于标签集中每一条皮肤三参数对应的 PBPK 输出中的 $A_{plasma}(t)$, 将其图像中的三个极值点设置为采样点, 记录这三个点的位置. 另外, 为了捕捉到更多曲线变化剧烈处的特征信息, 将满足 $\underset{0 < t < \text{极小值点}}{\operatorname{argmin}} |A_{plasma}(t) - 0.5 \cdot MAX_{plasma}|$ 的两个半衰期点也设置为采样点, 记录这两个点的位置. 将标签集对应的所有血药含量曲线的五个采样点都绘制在同一图像中, 如图3.4所示:

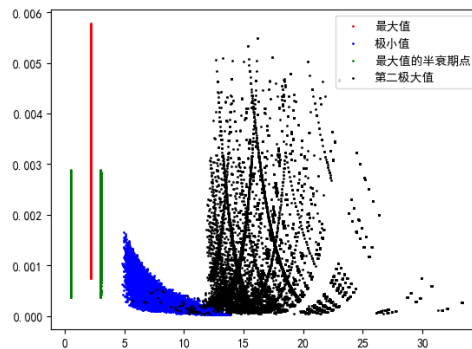


图 3.4: 所有 $A_{plasma}(t)$ 曲线的五个采样点

可以看到最大值点与两个半衰期点的位置几乎是固定的, 极小值点与第二极大值点几乎都在 $30h$ 前出现. 对于五种采样点, 计算图3.4中每种采样点位置的平均值, 将得到的五个平均值作为全局的固定采样点, 它们分别为 $(0.49, 2.165, 3.005, 8.22, 18.255)$.

对于 $A_{urinebps}(t)$ 与 $A_{urinebpsg}(t)$, 由于它们是浓度累计曲线 (如图3.2b), 难以像血浆 BPS 含量曲线一样从曲线形状来确定较少的采样点. 由于 $A_{urinebps}(t)$ 是数值求解的结果, 本质上是一个离散的数组, 故可以对其做一阶差分, 得到某个 $A_{urinebps}(t)$ 的一阶差分曲线如图3.5所示:

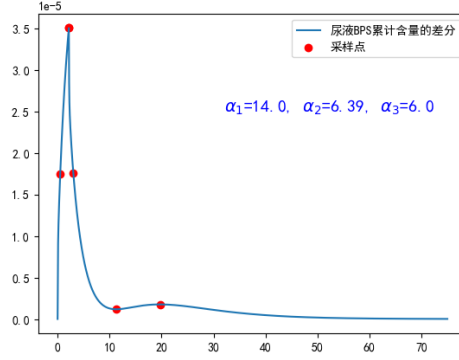
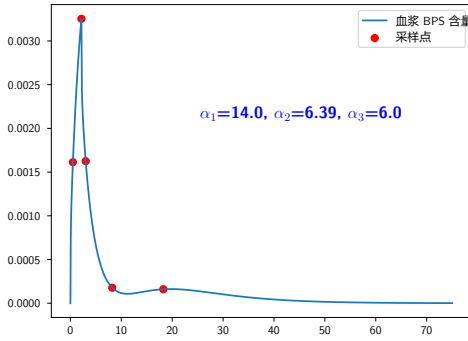
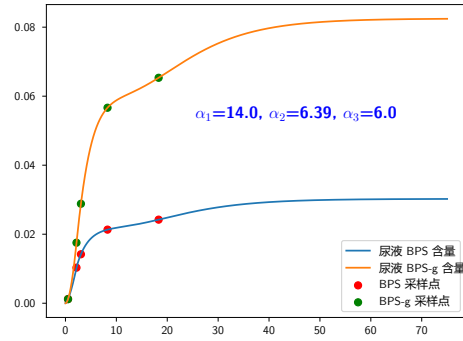


图 3.5: $\vec{\alpha} = (14, 6.39, 6)$ 时 $A_{urinebps}(t)$ 的一阶差分曲线

可以看到 $A_{urinebps}(t)$ 的一阶差分曲线有类似于 $A_{plasma}(t)$ 曲线的图像模式, 利用得到血浆曲线的五个全局固定采样点的方式可以得到两条尿液化学品累计含量曲线的五个全局固定采样点, 它们分别为 $(0.505, 2.165, 3.04, 8.265, 13.305)$. 图3.6展示了三条含量曲线的五个固定采样点的位置.



(a) $A_{plasma}(t)$ 与 5 点采样点



(b) $A_{urinebps}(t), A_{urinebpsg}(t)$ 与 5 点采样点

图 3.6: $\vec{\alpha} = (14, 6.39, 6)$ 时 PBPK 模型输出的部分曲线与采样点的图像 B

确定了血浆曲线与尿液曲线的全局固定采样时间节点后, 便可以确定特征集. 一条标签对应的三条曲线在固定时间节点处共采样了 15 个化学品含量数据, 一条特征即为含有 15 个元素的一维数组, 对应的标签为 PBPK 模型的输入皮肤三参数组, 整个特征集为形状是 47743×15 的二维数组.

3.3 参数反演神经网络模型的架构

为了使网络模型在多全连接层的情况下的性能退化问题, 本文的神经网络模型使用了残差连接的方法, 具体实现手段为引入了一个残差网络模块. 该模块由六层构成, 第一层和第四层为全连接层, 第二层和第五层为标准化层, 第三层和第六层为激活函数层, 使用的激活函数为 $Relu()$. 图片3.7展示了该模块的架构, 其中每一层的输入和输出的二维数组的形状都是 $batchsize \times hiddensize$, 其中 $batchsize$ 是一个批次的数据条数, $hiddensize$ 是可自定义的隐藏层神经元数量.

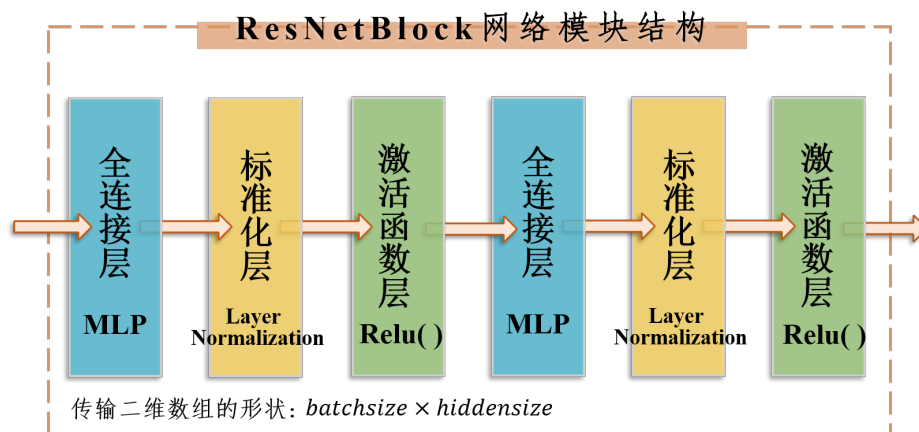


图 3.7: 残差网络模块的架构

该模块的最终输出结果是第六层的输出与第一层的输入的和, 相对于第六层输出的是期望输出和真实输出之间的残差, 这样的设定使得网络可以学习到残差映射, 同时保留了原始的输入并将其直接传输到了该模块的输出^[17-18]. 残差网络中的跳跃连接能够缓解深层网络下的性能退化与优化困难问题^[18]. 标准化层使用了层标准化手段 (layer-normalization), 使得输入数据在输出该层时能够具有零均值与单位方差的性质, 同时该标准化手段并不依赖于 $batchsize$ 的大小^[19]. 层标准化有效解决了深度网络情况下梯度爆炸与梯度消失的问题. 激活层的 $Relu()$ 的作用是对输入该层的数据逐元素取其正部. 它能够提高计算速率, 使得前向与反向传播的速度更快, 且保留正部而不是压缩数值的激活方式能够缓解梯度消失的问题^[20].

整个神经网络的架构串联拼接了 $hidden_nums$ 个残差网络模块, 并在串联组前后都添加了一层全连接网络, 输入端还额外添加了一个激活函数 $Relu()$ 层. 图片3.8展示了该模块的架构.

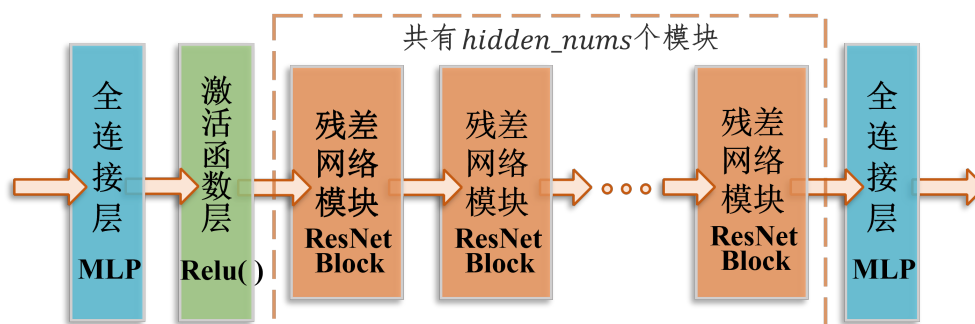


图 3.8: 本文神经网络模型的结构

网络的输入数据形状为 $batchsize \times inputsize$, 第一层全连接层的输出数据形状为 $batchsize \times hiddensize$, 后续层的输入输出至倒数第一层的输入数据的形状都是 $batchsize \times hiddensize$, 最后一层全连接层的输出数据形状为 $batchsize \times outputsize$.

对于该章的参数反演神经网络, 其网络模型架构的参数如下:

$$outputsize = 3, \quad hiddensize = 30, \quad hidden_nums = 3.$$

整个网络的输出即为反演后的皮肤三参数组, 中间隐藏层的神经元数量为 30 个, 残差网络模块的串联数量为 3 个. 同时, 根据特征集的不同, 本节共有两个参数反演神经网络, 两个网络模型的架构只在 $inputsize$ 的取值上有差异. “血药含量采样 28 点 & 尿药含量采样 15 点”特征集对应的 $inputsize = 58$, “血药含量采样 5 点 & 尿药含量采样 5 点”特征集对应的 $inputsize = 15$.

3.4 参数反演神经网络模型的训练

本章中的两个参数反演神经网络模型的训练部分是相同的, 故不作区分.

3.4.1 数据集的预处理, 划分与训练流程

在训练开始前, 将特征集与标签集分别进行标准化处理, 使作为数据集的矩阵的每一列具有零均值, 单位标准差的性质. 数据集的标准化处理可以增强模型的稳定性, 保持梯度的一致性与稳健性, 减少梯度消失或爆炸的风险^[21]. 同时, 标准化处理可以减小数据条目之间的数值差距, 使不同尺度的数据具有类似的分布特性, 令网络训练更专注于反演模型拟合这一主要任务, 而不是专注于输入数据的数值本身. 因而, 在处理未见到的数据时, 网络也不会因为陌生的数值而给出不佳的表现, 故标准化处理还可以提升网络的泛化能力^[22].

数据集预处理后, 根据训练需求被随机划分为训练集, 验证集, 测试集三个部分, 三个部分的条目比例为 8 : 1 : 1, 特征集与标签集被同步地划分, 划分后依然保持条目一一对应的关系.

在一轮训练中, 整个训练集被分为若干小批量数据, 在一轮训练内, 每个小批量数据都会前后被代入至未完善模型中, 计算损失函数关于网络模型内各系数的梯度信息, 并根据梯度信息使用优化算法来更新网络模型中的系数^[23]. 利用所有训练集小批量完成了一轮网络模型系数更新后, 将整个验证集分批次地代入至当前的网络模型中, 观测其对应的损失函数^[23]. 在验证期间, 引入早停技术 (early stopping) 中的耐心值设定: 记录当前有着验证集最优损失的训练轮次, 若接下来连续 $patience = 8$ 个训练轮内验证集上的损失函数没有小于验证集最优损失, 则停止训练并储存当前有着最优验证集损失的网络模型的全部系数, 否则继续下一轮的训练 (网络模型系数更新)^[23]. 训练结束后, 调用已储存的网络模型, 将验证集内的数据代入至网络模型计算损失函数, 并以此为评价网络模型优劣的指标之一.

3.4.2 损失函数与优化器

损失函数的设置利用了对偶学习 (Dual Learning) 的手段, 具体为 $\|\vec{\alpha} - \vec{\alpha}^*\|^2$ 与 $\|f(\vec{\alpha}) - \tilde{f}(\vec{\alpha}^*)\|^2$ 的线性组合. 其中 $\vec{\alpha}$ 为标签皮肤三参数组, $\vec{\alpha}^*$ 为网络模型输出的反演得到的皮肤三参数组, f 为公式 (3.1) 中的 PBPK 模型的求解模型, \tilde{f} 为拟合了 PBPK 模型正向求解的神经网络模型. 由于共有三种化学品含量曲线 $A_{plasma}(t)$, $A_{urinebps}(t)$, $A_{urinebpsg}(t)$, 一个参数反演神经网络共对应三个正向

拟合神经网络. 每个正向拟合神经网络的输入数据为参数反演网络输出的皮肤三参数组 $\vec{\alpha}^*$, 输出的是拟合后的对应于这组皮肤三参数组的某种化学品浓度曲线的采样点, 在下一节会详细介绍对偶学习与 PBPK 模型拟合神经网络模型的内容. 以“血药含量采样 28 点 & 尿药含量采样 15 点”对应的参数反演网络为例, 其损失函数中的 $\|f(\vec{\alpha}) - \tilde{f}(\vec{\alpha}^*)\|^2$ 具体形式为

$(MSE_{plasma} + MSE_{urinebps} + MSE_{urinebpsg}) \times \frac{1}{3}$, 其中三个 MSE 分别代表特征集中血药含量曲线采样点与正向拟合神经网络输出的血药含量曲线采样点之间的均方误差, 特征集中尿液 BPS 累计含量曲线采样点与正向拟合神经网络输出的尿液 BPS 累计含量曲线采样点之间的均方误差以及特征集中尿液 BPS-g 累计含量曲线采样点与正向拟合神经网络输出的尿液 BPS-g 累计含量曲线采样点之间的均方误差.

优化器使用了 torch.optim 中的 Adam() 函数, 即 Adam 算法. Adam 优化算法中的下降步长需要使用当前梯度信息的一阶矩估计 (梯度信息的指数平滑) 与二阶矩估计 (梯度信息的逐项平方的指数平滑)^[24]. 矩估计在经过修正后 (迭代步较小时的矩估计将会被放大), 代入至系数更新公式 $\Delta\theta = -\epsilon \frac{\hat{s}}{\sqrt{\hat{r} + \delta}}$, 其中 \hat{s} 为修正后的一阶矩估计, \hat{r} 为修正后的二阶矩估计, δ 为用于数值稳定的小常数^[24]. Adam 算法是一种高效的自适应学习率优化算法, 它可以利用迭代内的历史梯度信息来调整网络模型内每个系数的学习率, 使得模型系数的更新过程更加灵活与快速^[24]. 同时该算法不易陷入局部最优解, 可以很好地处理非凸问题^[25].

3.5 对偶学习与 PBPK 模型拟合神经网络模型

3.5.1 对偶学习的引入背景

在神经网络的训练过程中, 需要设置损失函数, 并以减少损失函数的数值为目标来迭代网络模型中的系数. 对于参数反演的问题, 见公式 (3.1), 损失函数应设置为 $\|f(\vec{\alpha}) - f(\vec{\alpha}^*)\|^2$, 即原本的皮肤三参数组代入至 PBPK 模型求解得到的含量曲线采样点与网络模型输出的反演皮肤三参数组代入至 PBPK 模型求解得到的含量曲线采样点之间的平均平方误差. 但由于计算资源问题, 较为庞大的 PBPK 模型的求解模型难以植入神经网络模型并在反向传播中正常运作. 故损失函数设置为 $\|\vec{\alpha} - \vec{\alpha}^*\|^2$, 即原本的皮肤三参数组与网络模型输出的反演皮肤三参数组之间的平均平方误差. 但这样的损失函数在训练过程中缺乏对神经网络模型的引导, 此处引入对偶学习, 训练一个能够拟合函数 f 的正向神经网络模型 \tilde{f} , 与参数反演神经网络形成对偶关系. 同时, 将 $\|f(\vec{\alpha}) - \tilde{f}(\vec{\alpha}^*)\|^2$ 作为损失函数内的一个正则化项, 与 $\|\vec{\alpha} - \vec{\alpha}^*\|^2$ 一项共同引导参数反演神经网络去学习 $\vec{\alpha}$ 与 $f(\vec{\alpha})$ 之间的映射关系. 其中 $f(\vec{\alpha})$ 已知, 即为参数反演神经网络的特征集, $\tilde{f}(\vec{\alpha}^*)$ 是参数反演神经网络输出的皮肤三参数组代入至正向拟合神经网络输出的结果, 含义为拟合得到的化学品含量曲线上的采样点.

3.5.2 正向拟合神经网络模型的数据集结构与构建

由于与参数反演神经网络是对偶关系, 本章中的正向拟合神经网络也使用固定时间采样节点来构建数据集, 同时也沿用参数反演神经网络数据集内的数据. 由于采样时间节点的不同, 本章共有两种参数反演神经网络, 相应地也有两类正向拟合神经网络. 并且, 为了更好地拟合 PBPK 模型输出的三种含量曲线 ($A_{plasma}(t)$, $A_{urinebps}(t)$, $A_{urinebpsg}(t)$), 对于每种时间采样点, 本节将分别训练三个正向拟合神经网络, 分别拟合血浆 BPS 含量曲线的采样点, 尿液 BPS 累计含量曲线的采样点以及尿液 BPS-g 累计含量曲线的采样点. 故正向拟合神经网络的特征集为皮肤三参数组 $\vec{\alpha}$ 的集合, 对应

的标签集为 $\vec{\alpha}$ 代入至 PBPK 模型求解得到的某种含量曲线的采样点 $(A(t_1), A(t_2), \dots, A(t_n))$ 的集合.

对于任意一种时间采样点, 需要训练的三个正向拟合神经网络的特征集都与相同时间采样点情形下的参数反演神经网络的标签集相同, 其形状为 47743×3 . 三个正向拟合神经网络的标签集是相同时间采样点情形下的参数反演神经网络的特征集的切片. 例如, 在血药含量曲线采样 28 点的情况下, 用于拟合 $A_{plasma}(t)$ 曲线的采样点的正向拟合神经网络的标签集是参数反演神经网络的特征集的前 28 列, 形状即为 47743×28 .

3.5.3 正向拟合神经网络模型的网络结构与训练

本节共有六个正向拟合神经网络, 对应两种时间采样节点与三种含量曲线 ($2 \times 3 = 6$). 对于每个正向拟合神经网络, 其架构与图3.8所呈现的相同, 架构中的参数相应地设置为:

$$inputsize = 3, \quad hiddensize = 30, \quad hidden_nums = 3.$$

整个网络的输入即为包含 47743 条皮肤三参数的特征集, 中间隐藏层的神经元数量为 30 个, 残差网络模块的串联数量为 3 个. 同时, 根据标签集的不同, 本小节内的六个网络模型的架构只在 $outputsize$ 的取值上有差异. 例如, 用于拟合“血药含量采样 28 点”的正向拟合神经网络对应的 $outputsize = 28$.

正向拟合神经网络模型训练时的损失函数为网络输出与对应标签之间的均方误差, 余下的训练方法与流程与3.4节中参数反演神经网络模型的训练完全相同, 不再介绍.

3.6 参数反演神经网络模型的效果展示

评价神经网络模型需要使用网络训练过程中未输入过的“陌生”数据, 以此来检查网络模型的泛化能力. 原始数据集划分出的测试集即是可用于测试网络模型的陌生数据.

评价指标说明

参数反演网络的输入为特征集内的 58 个元素或 15 个元素的含量曲线采样点, 输出为反演得到的皮肤三参数组. 为了评价网络模型的优劣, 需要将网络模型未见过的陌生数据输入至网络中, 并观察输出与陌生数据的标签之间的差异. 陌生数据 (测试数据) 也由特征和标签组成, 网络的前两个评价指标就是测试数据标签中的皮肤三参数组与网络输出的皮肤三参数组之间的均方误差 $MSE_{\vec{\alpha}}$ 与平均相对误差 $MRE_{\vec{\alpha}}$. 陌生数据的特征代表了三种化学品含量曲线上的 58 个采样点或 15 个采样点, 记为特征集化学品含量信息. 为了与之对应, 将网络输出的皮肤三参数组代入至 PBPK 模型的求解模型中, 在得到的三种化学品含量曲线上在固定时间节点上采样, 得到 58 个或 15 个元素的化学品含量信息, 记为网络输出的化学品含量信息. 将特征集化学品含量信息与网络输出的化学品含量信息之间的均方误差 $MSE_{concentration}$ 与平均相对误差 $MRE_{concentration}$ 作为另外两个评价指标, 分别简记为 MSE_C 与 MRE_C . 另外, 对于单条特征集化学品含量信息与对应的网络输出的化学品含量信息, 计算它们之间的决定系数^[26]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}.$$

其中 n 是一条化学品含量信息的元素数, y_i 是一条特征集化学品含量信息的第 i 个元素, \bar{y} 是一条特征集化学品含量信息的平均值, \hat{y}_i 是一条网络输出的化学品含量信息的第 i 个元素. R^2 原本被用于度量线性回归模型的预测数据与真实数据之间的差距^[26], 但也可应用于神经网络模型的预测数据与真实数据的差距衡量, 孙茜童等人和李爱民等人^[27-28]在评价参数反演功能的神经网络时都使用了 R^2 作为主要评价指标.

本文共使用五种评价指标: $MSE_{\bar{\alpha}}$, $MRE_{\bar{\alpha}}$, MSE_C , MRE_C , R^2 .

扩充测试集

由于测试集和神经网络训练时所使用的数据集是同源的, 欲考察神经网络在测试集之外的表现情况, 设定扩充测试集并将其应用于参数反演神经网络模型. 扩充测试集的构建与3.2节的数据集构建类似, 首先利用皮肤三参数组内各参数的截断正态分布分别对三个参数进行独立的采样, 再取三个参数采样集合的笛卡尔积, 得到标签集. 但为了测试网络模型的在更广的参数空间内的泛化能力, 对此处的截断正态分布的标准差和左右截断点都做出了改动, 使得依更改后的截断正态分布采样能更大概率地得到原数据集中出现频率较低的数值区间内的参数值. 原概率分布的参数为:

$$\begin{vmatrix} mean_1 = 17.28 & std_1 = 5.184 & lb_1 = 7.12 & ub_1 = 27.44 \\ mean_2 = 6.39 & std_2 = 1.917 & lb_2 = 2.63 & ub_2 = 10.1 \\ mean_3 = 5.7 & std_3 = 1.71 & lb_3 = 2.35 & ub_3 = 9.05 \end{vmatrix}$$

更改后的概率分布的参数如下表所示, 其中平均值 $mean_i$ 未变, 三个分布的标准差 std_i 都增大了, 后两个参数增大的幅度额外大, 三个参数的左右截断点之间的距离增大了.

$$\begin{vmatrix} mean_1 = 17.28 & std_1 = 8 & lb_1 = 5.78 & ub_1 = 28.78 \\ mean_2 = 6.39 & std_2 = 5 & lb_2 = 1.39 & ub_2 = 12.89 \\ mean_3 = 5.7 & std_3 = 5 & lb_3 = 0.7 & ub_3 = 11.7 \end{vmatrix}$$

对于每个参数, 依其更改后的截断正态分布采样 10 次, 取笛卡尔积后得到 $10 \times 10 \times 10 = 1000$ 组皮肤三参数组, 这是扩充测试集的标签集. 接下来将标签集中的皮肤三参数组代入至 PBPK 模型求解, 得到的三种化学品含量曲线 $A_{plasma}(t)$, $A_{urinebps}(t)$, $A_{urinebpsg}(t)$ 按照 3.2节中的两种固定时间采样点分别采样为两个特征集.

参数反演神经网络 A: 特征集每行包含 58 个含量信息

测试集类型	$MSE_{\bar{\alpha}}$	$MRE_{\bar{\alpha}}$	MSE_C	MRE_C	R^2
原始测试集	0.26	2.30%	5.54×10^{-6}	13.62%	99.95%
扩充测试集	0.76	2.20%	2.51×10^{-4}	39.45%	98.38%

表 3.1: 参数反演神经网络 A 在测试集上的评价指标

如表3.1所示, 使用了 58 个含量信息的特征集的参数反演神经网络在原始测试集和数值较为极端的扩充测试集上都得到了很好的效果, 其中 R^2 都处在 98% 以上, $MRE_{\bar{\alpha}}$ 都在 2.5% 以内. MRE_C 的数值较大, 原因可能是在计算化学品含量信息之间的相对误差时, 某些含量信息的数值极

小 (数量级为 10^{-9}), 细微的绝对数值差别都会导致很大的相对误差. 总的来说, 参数反演神经网络在原始测试集上的效果更好, 因为原始测试集虽然为陌生数据, 但与训练所用数据同源, 参数分布情况相近. 但在参数分布较极端的扩充测试集上, 网络模型 A 也显示出了不错的泛化效果. 图3.9展示了 $\bar{\alpha} = 16.91, 10.14, 7.95$ 时参数反演神经网络 A 的效果, 其中的曲线为特征集化学品含量信息对应的未采样前的完整曲线, 散点为网络输出的化学品含量信息.

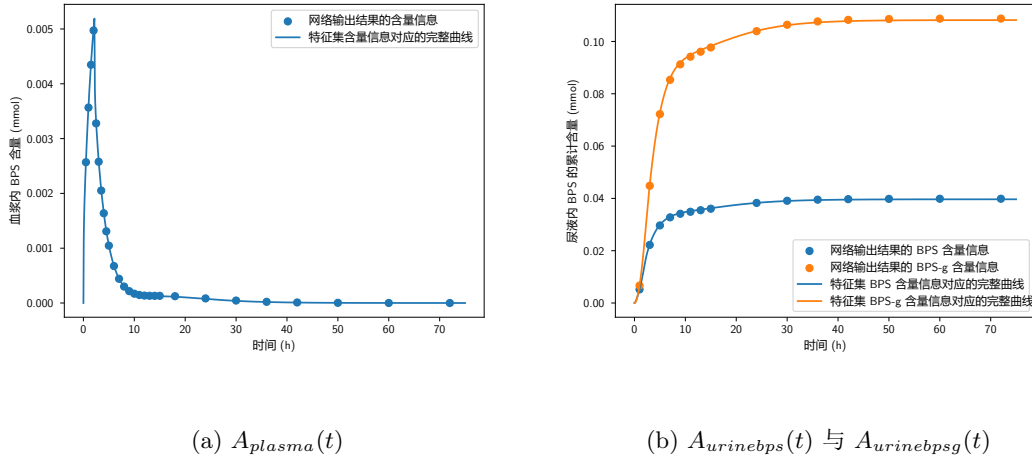


图 3.9: $\bar{\alpha} = 16.91, 10.14, 7.95$ 时参数反演神经网络 A 的输出可视化

参数反演神经网络 B: 特征集每行包含 15 个含量信息

测试集类型	$MSE_{\bar{\alpha}}$	$MRE_{\bar{\alpha}}$	MSE_C	MRE_C	R^2
原始测试集	0.17	1.82%	3.78×10^{-7}	1.39%	99.98%
扩充测试集	0.95	3.50%	1.52×10^{-6}	12.90%	93.91%

表 3.2: 参数反演神经网络 B 在测试集上的评价指标

如表3.2所示, 使用了 15 个含量信息的特征集的参数反演神经网络 B 在原始测试集得到了很好的效果, 甚至要优于使用了 58 个含量信息的特征集的参数反演神经网络 A 的表现. 在扩充测试集上, 由于特征集承载信息量不高, 参数反演神经网络 B 的表现不如参数反演神经网络 A 在扩充测试集上的表现, 同时相较于在原始测试集上的表现也有明显下滑, 但 R^2 仍然处在 94% 附近, $MRE_{\bar{\alpha}}$ 也有 3.50%. 另外, 参数反演神经网络 B 在两个测试集上的 MRE_C 明显要更小, 原因是由于采样时间点都在 30h 之前, 化学品含量数值的数量级并不十分小, 相对误差的计算未受很大影响. 总的来说, 参数反演神经网络 B 在两个测试集上的效果都不错, 在原始测试集上的效果与参数反演神经网络 A 几乎持平, 在参数分布较极端的扩充测试集上的效果比参数反演神经网络 A 稍差. 图3.10展示了 $\bar{\alpha} = 16.91, 10.14, 7.95$ 时参数反演神经网络 B 的效果, 其中的曲线为特征集化学品含量信息对应的未采样前的完整曲线, 散点为网络输出的化学品含量信息.

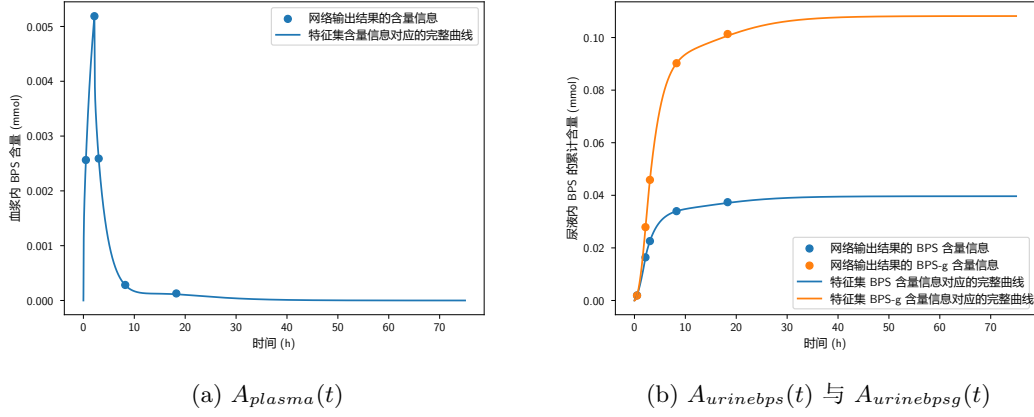


图 3.10: $\vec{\alpha} = (16.91, 10.14, 7.95)$ 时参数反演神经网络 B 的输出可视化

总结

根据在测试集上的评价指标可认为, 两个使用了不同形状的特征集进行训练的参数反演神经网络都胜任了拟合 PBPK 模型求解模型中的函数 f (见公式 (3.1)) 的反函数这一任务. 同时通过数值实验测试也验证了: 减少时间采样点的数量也能得到有效的参数反演结果. 这一结果可在深度学习数值测试意义上给药代动力学的实验提供了一种可能的理论基础: 在对模型进行参数校准时, 减少或精简对受试者采血或采尿的频次也可以得到与较多采样频次的情况相差并不大的参数校准结果, 以此来控制实验成本, 优化成本效益比, 获得更高的实验性价比.

第四章 PBPK 模型参数反演的神经网络方法：使用随机时间节点数据

本章聚焦于利用 Hu 等人^[1]提供的人体实验真实数据进行皮肤三参数组的参数校准，在第三章的基础之上在数据集中加入时间信息，利用对偶学习的方法，使用 PBPK 模型正向生成的带有时间信息的数据分别训练了拟合 PBPK 模型正向求解模型的神经网络以及 PBPK 模型参数反演的神经网络。本章的神经网络模型可以接受任意时间节点处的化学品含量信息。本章将讲解带有时间信息的数据集的结构与构造以及网络模型的结构，训练与效果，并利用参数反演神经网络与真实数据进行 PBPK 模型的参数校准。

4.1 参数校准与网络模型的基本介绍

4.2 Hu 等人^[1]的研究提供的人体实验真实数据与本文的参数反演神经网络模型基础架构适配的部分为某位受试者在 $0 \sim 75h$ 内在 14 个时间节点处的尿液内累计的 BPS 与 BPS-g 含量之和（尿液累计含量的具体定义详情见 3.1 节）。这些数据对应了 14 组的（排尿时间 t_i ，尿液累计化学品含量 A_i ）。在本章内，欲将这些数据作为某种参数反演神经网络的输入，输出反演得到的皮肤三参数组，以一新方法完成 Hu 等人^[1]研究中带平行分层皮肤仓室的 PBPK 模型的参数校准工作。

若利用第三章的使用固定时间信息的参数反演神经网络架构来做参数校准，需要将真实数据中的 14 个排尿时间作为固定的时间采样节点，建立单条含有 14 个化学品含量信息的特征集，并训练神经网络模型。但这种处理方式在时间节点的处理上缺乏灵活性，且真实的采样往往无法严格按照固定时间节点来执行。故本章在神经网络模型的数据集中引入时间信息，使得参数反演神经网络的输入中可加入时间点信息，这意味着不再需要固定的时间节点，在进行参数校准时可以同时输入真实数据中的排尿时间。为了使用真实数据内的全部 14 组数据，设置参数反演神经网络模型的输入为 $(A(t_1), \dots, A(t_{14}), t_1, \dots, t_{14})$ ，其中后 14 个元素为 $0 \sim 75h$ 内的 14 个时间点信息，前 14 个元素为尿液中 BPS 与 BPS-g 累计含量之和曲线（简称为尿液 BPS 类含量曲线）在这 14 个时间点上的采样。网络模型的输出依旧是反演得到的皮肤三参数组 $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ 。根据以上信息，可知使用时间信息数据的参数反演神经网络是在拟合如下的函数：

$$h : (\mathbb{R}^{14})^+ \times (0, 75)^{14} \rightarrow \mathbb{R}^3, \quad (A_1, \dots, A_{14}, t_1, \dots, t_{14}) \mapsto (\alpha_1, \alpha_2, \alpha_3). \quad (4.1)$$

由于使用对偶学习的手段，需要训练一个使用时间信息数据的 PBPK 模型的求解模型的正向拟合神经网络，这个神经网络是在拟合如下的与 PBPK 模型求解有关的函数：

$$g : (\mathbb{R}^3)^+ \times (0, 75) \rightarrow \mathbb{R}^+, \quad (\alpha_1, \alpha_2, \alpha_3, t) \mapsto A_{BPS}. \quad (4.2)$$

即对这个神经网络输入一个皮肤三参数组与一个时间点信息，可得到拟合的尿液 BPS 类含量曲线在该时间点的数值。

4.2 参数反演神经网络模型的数据集结构与构建

本章参数反演神经网络使用的整个数据集由大量数据组成, 设置数据条数为 N , 根据节的介绍, 整个特征集是一个形状为 $N \times 28$ 的二维数组, 与之对应的标签集是一个形状为 $N \times 3$ 的二维数组. 其中特征集由 14 个时间节点信息与 14 个对应时间点处的尿液 BPS 类含量曲线上的采样点构成, 称后者为尿液 BPS 类含量信息.

本章参数反演神经网络使用的标签集构造与小节中的标签集完全相同, 其中互不相同的皮肤三参数组共 47743 组. 由于要加入时间信息, 一条标签内的皮肤三参数组对应的特征是 14 个时间点信息与尿液 BPS 类含量曲线上的采样点. 为了使神经网络模型充分训练, 捕捉到时间信息在需要拟合的函数 g 中的作用, 需要尽可能使同一皮肤三参数组对应足够多的时间点信息, 故使一种皮肤三参数组对应 20 组 14 个点的时间信息, 意为皮肤三参数组代入至 PBPK 模型计算得到的尿液 BPS 类含量曲线共需采样 20 轮, 每轮共采样 14 个点并记录采样时间点信息, 其中每轮内的采样时间点并不相同且每轮采样得到的数据单独视为一个特征条目. 依然秉持标签集和特征集内条目一一对应的原则, 使标签集内的每一个条目都复制至 20 条, 故本章中参数反演神经网络模型使用的标签集一共 $47743 \times 20 = 954860$ 条, 标签集形状为 954860×3 .

对于标签集内 47743 组互不相同的皮肤三参数组内的每一组, 将其代入至 PBPK 模型的求解模型, 得到尿液 BPS 类含量曲线. 每条含量曲线需要采样 20 轮, 且每轮的采样时间点并不相同. 为了使采样时间点能契合真实数据的同时具有一定的随机性与普适性, 采用在 $0 \sim 75h$ 内随机抽取采样时间点的方法, 且将 $0 \sim 75h$ 划分为了许多小的抽取时间范围, 每个范围对应着固定的抽取个数, 以此来适配真实数据的时间点信息并避免抽取出的时间点在 $0 \sim 75h$ 内分布失衡的情况. 单轮的抽取规则如表 4.1 所示, 共抽取 14 个时间点.

抽取时间段	抽取个数
$0 \sim 6h$	2
$6 \sim 12h$	2
$12 \sim 18h$	2
$18 \sim 30h$	3
$30 \sim 46h$	2
$46 \sim 54h$	1
$54 \sim 64h$	1
$64 \sim 75h$	1

表 4.1: $0 \sim 75h$ 上的采样时间点抽取规则 (共 14 点)

每条尿液 BPS 类含量曲线的 20 轮内采样时间点都是完全不同的, 每条曲线需要抽取 20 轮采样时间点, 共 $47743 \times 20 = 954860$ 轮, 每一轮抽取都是互相独立的, 结果互不影响. 得到时间点信息后, 对尿液 BPS 类含量曲线采样, 得到的尿液 BPS 类含量信息与时间点信息为特征集的一个条目, 共 954860 条, 特征集的形状为 954860×28 . 图 4.1 展示了两个不同的尿液 BPS 类含量曲线与其随机抽取的 14 个采样点.

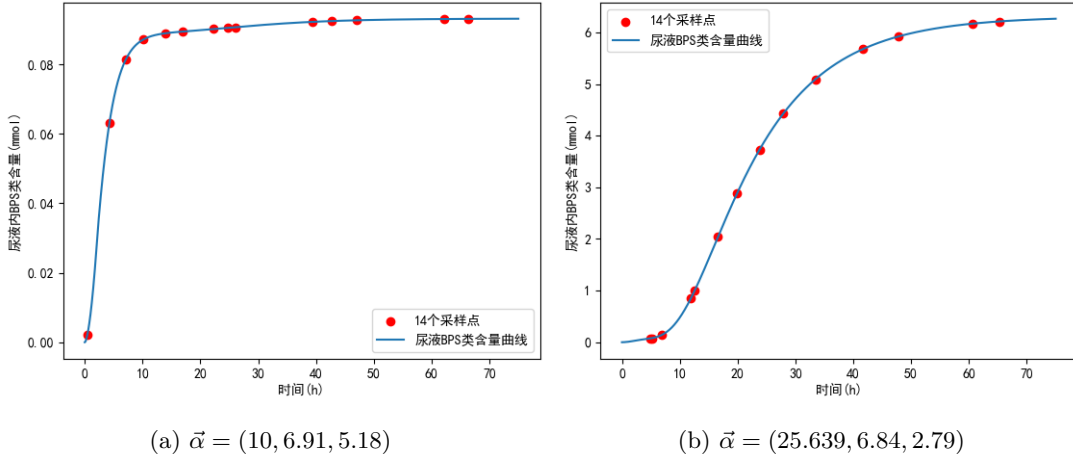


图 4.1: 两种尿液 BPS 类含量曲线与其随机采样点的示例

4.3 参数反演神经网络模型的网络结构与训练

本章的参数反演神经网络依旧使用了由全连接网络和残差网络模块构成的网络架构, 如图3.8. 另外本章的神经网络模型需要额外处理时间信息, 故增加隐藏层神经元的数量与残差网络模块的数量, 以加深网络, 进一步提高网络提取特征信息的能力. 架构中的参数相应地设置为:

$$inputs\ size = 28, \quad outputs\ size = 3, \quad hiddens\ size = 50, \quad hidden_nums = 5.$$

网络的输入为含有 14 个时间点信息与对应的尿液 BPS 类含量信息, 每条输入共 28 个元素, 网络的输出为反演得到的皮肤三参数组, 每台输出共 3 个元素, 中间隐藏层的神经元数量为 50 个, 残差网络模块的串联数量为 5 个.

本章的参数反演神经网络除损失函数外的训练流程与训练参数与3.4节中网络模型的训练相同, 此处重点讲解损失函数的选取. 损失函数的设置利用了对偶学习的方法, 单条数据对应的损失函数具体为 $\|\vec{\alpha} - \vec{\alpha}^*\|^2$ 与 $\sum_{i=1}^{14} [g(\vec{\alpha}, t_i) - \tilde{g}(\vec{\alpha}^*, t_i)]^2$ 的线性组合. 其中 $\vec{\alpha}$ 为标签皮肤三参数组, $\vec{\alpha}^*$ 为网络模型输出的反演得到的皮肤三参数组, g 为公式 (4.2) 中的与 PBPK 模型的正向求解有关的函数, \tilde{g} 为拟合了 g 的神经网络模型, 该网络模型的输入为一组皮肤三参数组与一个时间点信息, 输出为拟合的尿液 BPS 类含量曲线在该点的数值. 在下一节会详细介绍该正向拟合神经网络模型的内容.

4.4 可输入时间信息的 PBPK 模型拟合神经网络模型

作为参数反演神经网络训练所使用的损失函数的一部分, 本节中的神经网络模型拟合的是 PBPK 模型的正向求解模型内的一个部分所对应的函数 g (见公式4.2). 给定一组皮肤三参数组 $\vec{\alpha}$ 与 $0 \sim 75h$ 内的任意一个时间点 t_0 作为输入, 函数 g 可给出 $\vec{\alpha}$ 所对应的尿液内 BPS 类含量曲线在 t_0 处的取值.

正向拟合神经网络模型的数据集结构与构建

本节内正向拟合神经网络模型的输入为一个皮肤三参数组与一个时间点信息, 故特征集的形状为 $N \times 4$, 其中 N 为数据集的条目数. 本小节的特征集基于3.5.2小节中正向拟合神经网络模型的

特征集建立, 其中互不相同的皮肤三参数组共 47743 组. 为了使正向拟合神经网络模型充分学习到不同时间点信息的特征, 设定一条皮肤三参数组需要与 20 个不同的时间点信息组成 20 条含有四个元素的特征, 其中的 20 个时间点在 $0 \sim 75h$ 内随机抽取. 类似于4.2节中抽取时间点的方法, 将 $0 \sim 75h$ 分成许多时段, 并设置每个时段所需抽取的时间点数量, 这样的做法可以使时间点的分布尽量均匀, 以使特征集中含有充分的时间信息. 抽取时间点的规则如表4.2所示, 共抽取 20 个时间点.

抽取时间段	抽取个数	抽取时间段	抽取个数
$0 \sim 1h$	1	$12 \sim 16.5h$	1
$1 \sim 2h$	1	$16.5 \sim 21h$	1
$2 \sim 3h$	1	$21 \sim 25.5h$	1
$3 \sim 4h$	1	$25.5 \sim 30h$	1
$4 \sim 5h$	1	$30 \sim 37.5h$	1
$5 \sim 6h$	1	$37.5 \sim 45h$	1
$6 \sim 7h$	1	$45 \sim 52.5h$	1
$7 \sim 8h$	1	$52.5 \sim 60h$	1
$8 \sim 10h$	1	$60 \sim 67.5h$	1
$10 \sim 12h$	1	$67.5 \sim 75h$	1

表 4.2: $0 \sim 75h$ 上的采样时间点抽取规则 (共 20 点)

47743 个不同的皮肤三参数组, 每组都对应二十次时间点抽取, 共 $47743 \times 20 = 954860$ 次抽取, 所有抽取之间互相独立互不影响. 最后得到的特征集形状为 954860×4 . 标签集的构造较为简单, 每一条标签只由一个元素构成, 即是对应特征条目内的皮肤三参数组所对应的尿液 BPS 类含量曲线在特征条目内时间点上的采样值, 标签集的形状为 954860×1 .

正向拟合神经网络模型的网络结构与训练

本节中的正向拟合神经网络模型的架构和图3.8中相同, 其隐藏层神经元数量为 50 个, 串联残差网络模块有 5 个, 输入数据有 4 个元素, 输出元素有 1 个元素, 架构中的参数相应地设置为:

$$inputs\ size = 4, \quad outputs\ size = 1, \quad hiddens\ size = 50, \quad hidden_nums = 5.$$

本节的正向拟合神经网络除损失函数外的训练流程与训练参数与3.4节中网络模型的训练相同, 单条数据的损失函数为标签数据与输出数据之间的平方误差.

4.5 参数反演神经网络模型的效果展示

与3.6节相同, 本节将在数据集划分出的原始测试集与后续添加的扩充测试集上测试本章参数反演神经网络模型的效果.

评价指标说明

本章仍然使用 $MSE_{\bar{\alpha}}$, $MRE_{\bar{\alpha}}$, MSE_C , MRE_C , R^2 五种评价指标. 其中 $MSE_{\bar{\alpha}}$ 与 $MRE_{\bar{\alpha}}$ 分别为标签皮肤三参数组与网络输出的反演后的皮肤三参数组之间的均方误差与平均相对误差;

MSE_C , MRE_C 和 R^2 分别为特征集中前 14 个元素对应的尿液 BPS 类含量信息与网络输出的皮肤三参数组对应的相应时间点的尿液 BPS 类含量信息之间的均方误差, 平均相对误差与决定系数.

扩充测试集

扩充测试集内的皮肤三参数组在参数空间内的分布相较于原始测试集的标签集更加极端, 本章中扩充测试集的标签集的构建方法与 3.6 节中的完全相同, 都使用了截断正态分布采样的方法, 共得到 1000 条标签数据. 每条标签所对应的特征由两部分构成, 一部分是 14 个时间点信息, 另一部分是标签所对应的尿液 BPS 类含量曲线在这 14 个时间点上的采样值. 14 个时间点需要从 $0 \sim 75h$ 内随机抽取, 抽取规则与表 4.1 相同. 与原始数据集的构造不同, 扩充测试集中的每条标签只对应了一轮时间点的抽取. 故扩充测试集的特征有形状 1000×28 .

参数反演神经网络的效果

测试集类型	$MSE_{\bar{\alpha}}$	$MRE_{\bar{\alpha}}$	MSE_C	MRE_C	R^2
原始测试集	0.21	0.57%	1.16×10^{-5}	1.07%	99.75%
扩充测试集	0.90	1.71%	1.85×10^{-4}	2.11%	96.99%

表 4.3: 使用时间信息的参数反演神经网络在测试集上的评价指标

如表 4.3, 使用时间信息的参数反演神经网络模型在两个测试集上都得到了很好的效果, 无论是皮肤三参数组还是含量信息的平均相对误差, 都控制在了 2.2% 以内, 决定系数 R^2 也都在 96% 以上. 相对来说, 本章的参数反演神经网络比第三章不使用时间信息的参数反演神经网络的泛化效果更好, 我认为这可能与数据集的大小有关, 上一章的数据集条目数为 47743, 而本章的数据集条目数为 954860, 足足是前者的 20 倍. 且本章的网络模型使用了更深的神经网络架构, 这可能也是泛化能力更强的原因之一. 图 4.2 展示了测试集内标签为 $\bar{\alpha} = (19.75, 8.89, 3.73125)$ 的测试数据在本章参数反演神经网络中的输出含量信息与特征集含量曲线.

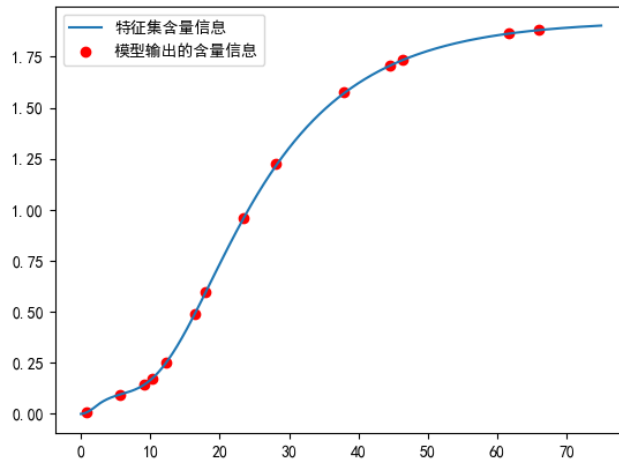


图 4.2: $\bar{\alpha} = (19.75, 8.89, 3.73)$ 时参数反演神经网络的输出可视化

4.6 基于真实实验数据^[1]的参数校准

将人体实验的真实数据, 14 个时间点信息与 14 个尿液 BPS 类含量信息, 输入至本章的参数反演神经网络中, 网络输出反演后的皮肤三参数组 $\vec{\alpha} = (14.9377, 6.1184, 3.7767)$, 作为校准后的皮肤三参数组. 将网络输出的皮肤三参数组代入至 PBPK 求解模型中, 得到一条尿液内 BPS 类含量曲线, 与前人校准得到的皮肤三参数组 $\vec{\alpha} = (17.28, 6.39, 5, 7)$ 对应的尿液内 BPS 类含量曲线和真实 BPS 类含量数据绘制在一起, 如图4.3.

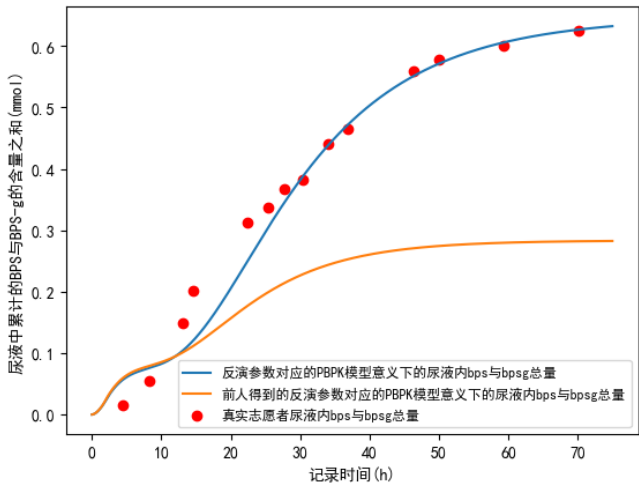


图 4.3: 皮肤三参数组校准结果展示

计算校准后的皮肤三参数组对应的 14 个时间点处的尿液 BPS 类含量信息与真实尿液 BPS 类含量信息之间的均方误差与平均相对误差, 得到的结果与两种方式的校准时间如表4.4所示. 可以看到本文的校准方式在速度和精度上都超过了前人的传统优化方法.

校准手段	MSE	MRE	校准时间
参数反演神经网络	1.17×10^{-3}	29.56%	0.001s
前人研究	6.21×10^{-1}	62.13%	>5min

表 4.4: 两种校准手段的指标展示

第五章 PBPK 模型的统计学分析

5.1 敏感性分析

敏感性分析 (Sensitivity Analysis, SA) 是一种用于评估模型输出关于模型输入的变化敏感度的分析手段^[29]. 在 Hu 等人^[1]的研究中, 将模型输出设置为 $A_{plasma}(t)$ 的曲线下面积 AUC 与最大值 C_{max} . 将模型的某种输入参数提高至原来的 1.1 倍, 或减少至原来的 0.9 倍, 并观察模型输出的变化幅度, 即可得到模型输出关于该种参数的敏感度.

在本文中, PBPK 模型只有皮肤三参数组 $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ 作为输入, 即 DSC, Pfo 与 u_1 . 三个参数的基准值设置为上一章中使用参数反演神经网络得到的校准值 $\vec{\alpha} = (14.9377, 6.1184, 3.7767)$, 并对每个参数分别进行提高至 1.1 倍与减少至 0.9 倍的操作 (即增减 10%), 记录 $A_{plasma}(t)$ 的曲线下面积 AUC 与最大值 C_{max} 相应的变化幅度. 将皮肤三参数组的基准值与更改值分别代入至 PBPK 模型的求解模型, 得到结果如表格 5.1 所示. 可以看出, AUC 对 DSC 变化的敏感性最高, 甚至百分比变化接近 50%, 其次是 u_1 , 对 Pfo 的敏感性很低. 另外, AUC 与 u_1 的相关性和与 DSC 或 Pfo 的相关性相反. 校准后的 DSC 与 u_1 似乎能使 C_{max} 处在极小值附近, 因为无论增还是减 DSC 或 u_1 , C_{max} 都会增加, 但增加的幅度不同. 相较于减小, 当 DSC 增加时, C_{max} 增加的幅度更大, 对于 u_1 也是如此. C_{max} 可能与 Pfo 是线性的关系, Pfo 增加 10% 时, C_{max} 也随之增加 10%, 当 Pfo 减少 10% 时, C_{max} 也随之减少 10%. 总得来说, 两个模型输出对皮肤三参数组的敏感性都较强, 说明了对皮肤三参数组进行校准的必要性.

参数处理情况	C_{max} 百分比变化	AUC 百分比变化
$DSC \times 1.1$	20.76%	46.65%
$DSC \times 0.9$	0.41%	-35.79%
$Pfo \times 1.1$	10.00%	1.23%
$Pfo \times 0.9$	-10.00%	-1.23%
$u_1 \times 1.1$	0.26%	-26.99%
$u_1 \times 0.9$	3.54%	37.26%

表 5.1: PBPK 模型输出对皮肤三参数组的敏感度

5.2 不确定性分析

不确定性分析是医学模型中常用的一个手段, 它的做法是利用蒙特卡洛模拟的方法依照某种概率分布抽取若干组模型的输入参数, 得到若干组模型输出, 之后将模型输出按从小到大的顺序排列, 并提取其中的 95 分数 (大于余下 95% 的数据的值) 与 50 分数 (即中位数)^[30]. 模型的不确定值被定义为 $uncertainty = \frac{P_{95}}{P_{50}}$. 该值越接近于 1, 说明模型输出对所选输入参数的不确定性越小, 该值大于 2 时, 说明模型的不确定性较高^[30].

Hu 等人^[1]在不确定性分析中使用了截断正态分布来依概率分布抽取模型的输入参数, 本节将对皮肤三参数组进行以概率分布的抽取, 并分别计算模型输出 $A_{plasma}(t)$ 的曲线下面积 AUC 与最大值 C_{max} 对三个参数 DSC , Pfo 与 u_1 的不确定性. 其中皮肤三参数组的截断正态分布的平均值使用本文的校准结果, 变异系数统一设置为 30%, 标准差为平均值与变异系数的乘积, 左截断点为 $lb = mean - 1.96 \times std$, 右截断点为 $ub = mean + 1.96 \times std$, 三个参数的截断正态分布的参数如下所示:

$$\left| \begin{array}{cccc} mean_1 = 14.9377 & std_1 = 4.48 & lb_1 = 6.16 & ub_1 = 23.72 \\ mean_2 = 6.1184 & std_2 = 1.84 & lb_2 = 2.512 & ub_2 = 9.7248 \\ mean_3 = 3.7767 & std_3 = 1.13 & lb_3 = 1.56 & ub_3 = 5.99 \end{array} \right|$$

对每个参数分别以概率分布抽取 10000 次, 分别计算 AUC 与 C_{max} 关于三个参数的不确定性, 得到结果如表 5.2 所示. 可以看到 C_{max} 关于 DSC 的不确定性最高, 超过了 3.5, 关于 u_1 的不确定性也很高, 关于 Pfo 的不确定性较低, 小于 1.5. AUC 关于 DSC 和 u_1 的不确定性都很高, 都超过了 3.5, 关于 Pfo 的不确定性数值极低, 接近于 1. 从不确定性分析可以看出, DSC 与 u_1 相对来说对模型输出的贡献更大, 更值得关注.

参数	C_{max} 的不确定性	AUC 的不确定性
DSC	3.68	3.81
Pfo	1.44	1.05
u_1	2.73	3.76

表 5.2: PBPK 模型输出关于皮肤三参数组的不确定性

第六章 总结

研究回顾

本文基于 Hu 等人^[1]的研究, 从双酚 S 的人体平行分层皮肤的 PBPK 模型的模型求解, 模型参数校准到模型不确定性分析完成了对该模型的二次解读与构建. 其中文章的主题与着重点在于利用神经网络模型来进行模型的参数校准. 在模型求解部分, 本文总结了 PBPK 模型的数学形式并使用 Python 代码构建了模型的数值求解模型. 在参数校准部分, 本文分别构建了两类参数反演网络, 分别是不包含时间节点信息的网络模型包含时间节点信息的网络模型, 网络的架构都是深度残差神经网络, 并使用对偶学习的方法完成了网络的训练. 利用包含时间节点信息的参数反演网络模型与 Hu 等人^[1]提供的人体实验数据完成了 PBPK 模型的参数校准. 最后, 利用校准好的参数完成了模型输出关于需校准的三个参数的敏感性分析与不确定性分析.

结论与讨论

模型求解

在模型求解部分, 本文利用 Hu 等人^[1]提供的 PBPK 模型的数学形式完成了求解代码的复现, 得到了与前人模型求解代码结果相同但速度超出其 100 倍的复现结果. 但 Hu 等人提供的模型在当某些皮肤三参数组作为模型输入的情况下, 会求解得到出现双峰图案的血药含量曲线, 这代表着 BPS 在血液内的含量快速下降后又出现了明显回升, 在大多数情况下, 双峰的图案是不正确的, 也许是 Hu 等人提供的与皮肤相关的人体参数并不准确, 或者是带平行分层皮肤仓室的 PBPK 模型本身并不完善.

参数校准与神经网络模型

为了更加快速且精准地得到参数反演的结果, 本文建立了神经网络模型, 输入离散的血液或尿液中 BPS 的含量信息, 网络模型会输出反演后的皮肤三参数组. 本质上, 参数反演神经网络是对 PBPK 模型的正向求解模型的反模型的拟合. 在原始测试集和扩充测试集上, 根据评价指标的数值可得出结论: 本文提出的神经网络模型显示出了很好的泛化能力. 另外, 在使用固定时间点的神经网络模型背景下, 本文在数值实验的意义上得到减少采样时间点也能得到良好反演效果的结论, 意味着在真实人体实验中, 可以减少对受试者的采血采尿的频次, 以降低实验成本, 优化成本效益比. 最后, 本文利用真实数据完成了参数校准, 校准后的参数与 Hu 等人^[1]的校准参数相比, 在真实数据拟合方面有着更优秀的表现, 且本文的参数校准方法速度更快.

统计学分析

本文的最后一部分基于校准后的参数, 对模型的两种输出 (血药含量曲线的曲线下面积与最大值) 进行了关于皮肤三参数组的敏感性分析和不确定性分析. 分析结果表明, DSC 与 u_1 是对模型

输出贡献较大的两个参数, 模型输出对这两个参数的不确定性和敏感性都很高, 这意味着这两个参数需要得到更进一步的模型解释与范围界定, 也可能意味着 Hu 等人^[4]提出的平行分层皮肤仓室模型的预测结果并不稳定, 存在较大风险, 需要改变模型结构或给出更合理的皮肤相关参数的取值.

结论

神经网络模型在 PBPK 模型框架下能够胜任参数反演或参数校准的任务. 若研究者做人体实验的目的为校准模型参数, 在使用神经网络模型作为校准工具的前提下, 可以考虑减少对受试者采集生物材料的频次, 以降低成本节约时间.

附录 A PBPK 模型物理量名称-含义对照表

模型中待求解变量

$C_{SCi}(t)$	暴露皮肤组织的深度为 $i \times \frac{T_{SC}}{10}$ 的角质层在 t 时刻的 BPS 浓度 ($nmol/cm^3$)
$A_{Fo}(t)$	暴露皮肤组织的毛囊在 t 时刻的 BPS 含量 ($nmol$)
$A_{well}(t)$	暴露皮肤组织的表皮储仓在 t 时刻的 BPS 含量 ($nmol$)
$A_{VE}(t)$	暴露皮肤组织的活性表皮在 t 时刻的 BPS 含量 ($nmol$)
$A_{ST}(t)$	胃部在 t 时刻的 BPS 含量 ($nmol$)
$A_{skin}(t)$	未暴露皮肤组织在 t 时刻的 BPS 含量 ($nmol$)
$A_{fat}(t)$	脂肪组织在 t 时刻的 BPS 含量 ($nmol$)
$A_{gonad}(t)$	性腺在 t 时刻的 BPS 含量 ($nmol$)
$A_{plasma}(t)$	血浆在 t 时刻的 BPS 含量 ($nmol$)
$A_{brain}(t)$	脑部在 t 时刻的 BPS 含量 ($nmol$)
$A_{rich}(t)$	血流丰富组织在 t 时刻的 BPS 含量 ($nmol$)
$A_{slow}(t)$	血流缓慢组织在 t 时刻的 BPS 含量 ($nmol$)
$A_{GIBPSg}(t)$	胃肠道在 t 时刻的 BPS-g 含量 ($nmol$)
$A_{GIBPSs}(t)$	胃肠道在 t 时刻的 BPS-s 含量 ($nmol$)
$A_{SI}(t)$	小肠在 t 时刻的 BPS 含量 ($nmol$)
$A_{liver}(t)$	肝脏在 t 时刻的 BPS 含量 ($nmol$)
$A_{BPSg_delay}(t)$	发生了肝肠循环的 BPS-g 的量/小肠在 t 时刻的 BPS-g 含量 ($nmol$)
$A_{BPSs_delay}(t)$	发生了肝肠循环的 BPS-s 的量/小肠在 t 时刻的 BPS-s 含量 ($nmol$)
$A_{BPSg}(t)$	整个机体在 t 时刻的 BPS-g 含量 ($nmol$)
$A_{BPSs}(t)$	整个机体在 t 时刻的 BPS-s 含量 ($nmol$)
$A_{urinebps}(t)$	人体排出的尿液在 t 时刻累计的 BPS 含量 ($nmol$)
$A_{urinebpsg}(t)$	人体排出的尿液在 t 时刻累计的 BPS-g 含量 ($nmol$)

其他变量

$\varphi(x, t)$	暴露皮肤组织的角质层深度 x 处在 t 时刻的 BPS 浓度 ($nmol/cm^3$),
$f_1(t)$	皮肤接触外源 BPS 的量 ($nmol$), 当 $t > Time_{add}$ 时, $f_1(t) = 0$
$ON(t)$	布尔值, 当 $t \leq Time_{expose}$ 时, $ON(t) = 1$, 皮肤处于 BPS 暴露状态; 当 $t > Time_{expose}$ 时, $ON(t) = 0$, 皮肤处于未暴露状态
$f_2(t)$	口服外源 BPS 的量 ($nmol$), 本文中不考虑口服情况, $f_2(t) = 0$

待反演的参数

DSC	角质层中的有效扩散系数 (cm^2/h)
u_1	由脱屑而向皮肤表面转移的速度 (cm/min)
Pfo	毛囊的渗透系数 (cm/h)

理化数据中的常量

T_{SC}	角质层的深度 (um)
$SCDX$	角质层的深度的 $\frac{1}{10}(um)$
HSC_{well}	角质层和皮肤表皮储仓之间的分配系数
V_{well}	暴露皮肤组织的表面储仓体积 (L)
HSC_{VE}	角质层和活性表皮之间的分配系数
V_{TVE}	暴露皮肤组织的活性表皮层体积 (L)
$AEXP$	暴露皮肤组织面积 (dm^2)
$FEXP$	暴露皮肤组织中毛囊的面积分数
HFO_{well}	毛囊和皮肤表皮储仓之间的分配系数
V_{TFO}	暴露皮肤组织的毛囊体积 (L)
Q_{skin}	皮肤血液流速 (L/h)
p_{skin}	皮肤-血浆分配系数
BSA	人体皮肤表面积 (dm^2)
V_{plasma}	血浆体积 (L)
k_0	口服给药时 BPS 从胃进入肝脏的系数 (h^{-1})
k_1	口服给药时 BPS 从小肠进入肝脏的系数 (h^{-1})
ge	口服给药时 BPS 由胃转移至小肠的系数 (h^{-1})
V_{TSC}	暴露皮肤组织的角质层体积 (L)
V_{skin}	皮肤组织的体积 (L)
Q_{fat}	脂肪组织血液流速 (L/h)
V_{fat}	脂肪组织的体积 (L)
p_{fat}	脂肪-血浆分配系数
Q_{gonad}	性腺血液流速 (L/h)
V_{gonad}	性腺的体积 (L)
p_{gonad}	性腺-血浆分配系数
$K_{urinebps}$	BPS 尿液排泄参数 (L/h)
Q_c	心脏血液流速 (L/h)
Q_{brain}	脑部血液流速 (L/h)
V_{brain}	脑部的体积 (L)
p_{brain}	脑部-血浆分配系数
Q_{rich}	血流丰富组织血液流速 (L/h)
V_{rich}	血流丰富组织的体积 (L)

$prich$	血流丰富组织-血浆分配系数
Q_{slow}	血流缓慢组织血液流速 (L/h)
V_{slow}	血流缓慢组织的体积 (L)
$pslow$	血流缓慢组织-血浆分配系数
Q_{liver}	肝脏血液流速 (L/h)
V_{liver}	肝脏的体积 (L)
$pliver$	肝脏-血浆分配系数
kGI_{ing}	口服给药 BPS-g 从肠到血中的系数 (h^{-1})
kGI_{ins}	口服给药 BPS-s 从肠到血中的系数 (h^{-1})
$V_{maxgutg}$	肠道中 BPS 葡萄糖酸化的最大反应速度 ($nmol/h$)
$V_{maxguts}$	肠道中 BPS 硫酸盐化的最大反应速度 ($nmol/h$)
$enterocytes$	小肠体积 (L)
K_{mgutg}	肠道中 BPS 葡萄糖酸化的米氏常数 ($nmol$)
K_{mguts}	肠道中 BPS 硫酸盐化的米氏常数 ($nmol$)
K_{sigutg}	肠道中葡萄糖酸化结合底物抑制常数 ($nmol$)
$k_{enterobpsg}$	BPS-g 肝肠循环使得 BPS 发生循环的速率 (h^{-1})
$k_{enterobps s}$	BPS-s 肝肠循环使得 BPS 发生循环的速率 (h^{-1})
$K_{mliverg}$	肝脏中 BPS 葡萄糖酸化的米氏常数 ($nmol$)
$V_{maxliverg}$	肠道中 BPS 葡萄糖酸化的最大反应速度 ($nmol/h$)
$K_{mlivers}$	肝脏中 BPS 硫酸盐化的米氏常数 ($nmol$)
$V_{maxlivers}$	肠道中 BPS 硫酸盐化的最大反应速度 ($nmol/h$)
$met1g$	肝脏中 BPS-g 进入血液中的比例
$met1s$	肝脏中 BPS-s 进入血液中的比例
$met2g = 1 - met1g$	肝脏中 BPS-g 进入肝肠循环的比例
$met2s = 1 - met1s$	肝脏中 BPS-s 进入肝肠循环的比例
k_{entero}	BPS-g 肝肠循环的速率 (h^{-1})
k_{4IV}	BPS-g 在肝肠循环中的粪便消除系数 (h^{-1})
$K_{urinebpsg}$	BPS-g 的尿液排泄参数 (L/h)
V_{bodyg}	参与 BPS-g 分布的组织体积 (L)
$K_{urinebps s}$	BPS-s 的尿液排泄参数 (L/h)
$V_{body s}$	参与 BPS-s 分布的组织体积 (L)

实验设置的常量

$Time_{add} = \frac{1}{6}h$	手指皮肤触摸热敏纸 (外源 BPS) 的时间 (h)
$Time_{expose} = \frac{13}{6}h$	手指皮肤表皮储仓内 BPS 含量大于 0 的时间 (h)

参考文献

- [1] HU M, ZHANG Z, ZHANG Y, et al. Development of human dermal pbpk models for the bisphenols bpa, bps, bpf, and bpaf with parallel-layered skin compartment: Basing on dermal administration studies in humans[J/OL]. Science of The Total Environment, 2023, 868: 161639. <https://www.sciencedirect.com/science/article/pii/S0048969723002541>. DOI: <https://doi.org/10.1016/j.scitotenv.2023.161639>.
- [2] VANDENBERG L N, HAUSER R, MARCUS M, et al. Human exposure to bisphenol a (bpa) [J/OL]. Reproductive Toxicology, 2007, 24(2): 139-177. <https://www.sciencedirect.com/science/article/pii/S0890623807002377>. DOI: <https://doi.org/10.1016/j.reprotox.2007.07.010>.
- [3] RUBIN B S. Bisphenol a: An endocrine disruptor with widespread exposure and multiple effects[J/OL]. The Journal of Steroid Biochemistry and Molecular Biology, 2011, 127(1): 27-34. <https://www.sciencedirect.com/science/article/pii/S0960076011001063>. DOI: <https://doi.org/10.1016/j.jsbmb.2011.05.002>.
- [4] 张益宁周颖. 双酚 A 的生理毒代动力学 (PBTK) 模型构建研究[J]. 食品与营养科学, 2021.
- [5] CORBEL T, GAYRARD V, PUEL S, et al. Bidirectional placental transfer of bisphenol a and its main metabolite, bisphenol a-glucuronide, in the isolated perfused human placenta[J/OL]. Reproductive Toxicology, 2014, 47: 51-58. <https://www.sciencedirect.com/science/article/pii/S0890623814001026>. DOI: <https://doi.org/10.1016/j.reprotox.2014.06.001>.
- [6] 赫淑铭刘娴 傅建捷张爱茜 江桂斌. 双酚类化合物的生物代谢机理研究进展[J/OL]. 环境化学, 2024, 43: 711. <http://hjhx.rcees.ac.cn/article/id/648a6a10c59bc3243d453c96>. DOI: [10.7524/j.issn.0254-6108.2023041401](https://doi.org/10.7524/j.issn.0254-6108.2023041401).
- [7] PELCH K, WIGNALL J A, GOLDSTONE A E, et al. A scoping review of the health and toxicological activity of bisphenol a (bpa) structural analogues and functional alternatives [J/OL]. Toxicology, 2019, 424: 152235. <https://www.sciencedirect.com/science/article/pii/S0300483X18306668>. DOI: <https://doi.org/10.1016/j.tox.2019.06.006>.
- [8] SHUYING ZHANG J C, ZhongYu WANG. Application of physiologically based toxicokinetics models in risk assessment of chemicals[J/OL]. Chinese Science Bulletin, 2017, 62(35): 4139-4150. <http://www.sciengine.com/publisher/ScienceChinaPress/journal/ChineseScienceBulletin/62/35/10.1360/N972017-00886>. DOI: <https://doi.org/10.1360/N972017-00886>.
- [9] 孙晋都. 纳米银体内外动力学研究及生理毒物动力学模型 (PBTK) 构建[D]. 南京, 江苏: 东南大学, 2017.

- [10] YANG X, DOERGE D R, TEEGUARDEN J G, et al. Development of a physiologically based pharmacokinetic model for assessment of human exposure to bisphenol a[J/OL]. *Toxicology and Applied Pharmacology*, 2015, 289(3): 442-456. <https://www.sciencedirect.com/science/article/pii/S0041008X15301198>. DOI: <https://doi.org/10.1016/j.taap.2015.10.016>.
- [11] KARRER C, ROISS T, VON GOETZ N, et al. Physiologically based pharmacokinetic (pbpk) modeling of the bisphenols bpa, bps, bpf, and bpaf with new experimental metabolic parameters: Comparing the pharmacokinetic behavior of bpa with its substitutes[J/OL]. *Environmental Health Perspectives*, 2018, 126(7): 077002. <https://ehp.niehs.nih.gov/doi/abs/10.1289/EHP2739>.
- [12] YINGZHOU8. Pbk[EB/OL]. 2022. <https://github.com/YingZhou8/PBK>.
- [13] KHMIRI I, Côté J, MANTHA M, et al. Toxicokinetics of bisphenol-s and its glucuronide in plasma and urine following oral and dermal exposure in volunteers for the interpretation of biomonitoring data[J/OL]. *Environment International*, 2020, 138: 105644. <https://www.sciencedirect.com/science/article/pii/S0160412019349967>. DOI: <https://doi.org/10.1016/j.envint.2020.105644>.
- [14] HINDMARSH A C. Odepack: A systematized collection of ode solvers[J]. *Scientific Computing*, 1983.
- [15] GARCKE J. Sparse grids in a nutshell[C/OL]//2012. <https://api.semanticscholar.org/CorpusID:33937930>. DOI: 10.1007/978-3-642-31703-3_3.
- [16] BUNGARTZ H J. Sparse grids[J/OL]. *Acta Numerica*, 2004, 13: 147 - 269. <https://api.semanticscholar.org/CorpusID:229170630>.
- [17] ZAEEMZADEH A, RAHNAVARD N, SHAH M. Norm-preservation: Why residual networks can become extremely deep?[J/OL]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(11): 3980-3990. DOI: 10.1109/TPAMI.2020.2990339.
- [18] LIU F, REN X, ZHANG Z, et al. Rethinking skip connection with layer normalization[C/OL]//SCOTT D, BEL N, ZONG C. *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020: 3586-3598. <https://aclanthology.org/2020.coling-main.320>. DOI: 10.18653/v1/2020.coling-main.320.
- [19] BA J, KIROUS J R, HINTON G E. Layer normalization: abs/1607.06450[A/OL]. 2016. <https://api.semanticscholar.org/CorpusID:8236317>.
- [20] LIU C, HUI L. Relu soothes the ntk condition number and accelerates optimization for wide neural networks[A]. 2023. arXiv: 2305.08813.
- [21] SHANKER M, HU M, HUNG M. Effect of data standardization on neural network training[J/OL]. *Omega*, 1996, 24(4): 385-397. <https://www.sciencedirect.com/science/article/pii/0305048396000102>. DOI: [https://doi.org/10.1016/0305-0483\(96\)00010-2](https://doi.org/10.1016/0305-0483(96)00010-2).

- [22] HAMANN I M, HERZFELD U C. On the effects of preanalysis standardization[J/OL]. The Journal of Geology, 1991, 99(4): 621-631[2024-05-10]. <http://www.jstor.org/stable/30065006>.
- [23] PETERSON G E. Foundation for neural network verification and validation[C/OL]//RUCK D W. Science of Artificial Neural Networks II: Vol. 1966. SPIE, 1993: 196 - 207. <https://doi.org/10.1117/12.152651>.
- [24] KINGMA D P, BA J. Adam: A method for stochastic optimization[J/OL]. CoRR, 2014, abs/1412.6980. <https://api.semanticscholar.org/CorpusID:6628106>.
- [25] BAI J, REN Y, ZHANG J. Bgadam: Boosting based genetic-evolutionary adam for neural network optimization[J/OL]. 2021 International Joint Conference on Neural Networks (IJCNN), 2019: 1-8. <https://api.semanticscholar.org/CorpusID:233738868>.
- [26] and 秦超英师义民, 许勇. 数理统计[M]. 4 版. 北京: 科学出版社, 2015.
- [27] 孙茜童, 付芸, 韩春晓, 等. 基于卷积神经网络的全球海洋叶绿素 a 浓度反演方法[J]. 光谱学与光谱分析, 2023, 43(2): 608-613.
- [28] 李爱民, 范猛, 秦光铎, 等. 卷积神经网络模型的遥感反演水质参数 COD[J]. 光谱学与光谱分析, 2023, 43(2): 651-656.
- [29] FREY H C, PATIL S R. Identification and review of sensitivity analysis methods[J]. Risk Analysis, 2002, 22(3): 553-578.
- [30] MCGRATTAN K, TOMAN B. Quantifying the predictive uncertainty of complex numerical models[J/OL]. Metrologia, 2011, 48(3): 173. <https://dx.doi.org/10.1088/0026-1394/48/3/011>.

致 谢

致谢