

doi: 10.7690/bgzdh.2026.03.016

基于长短期记忆网络和卡尔曼滤波的快速存取记录器数据降噪研究

杨军利, 杨沛达, 钱宇, 肖毅

(中国民用航空飞行学院飞行技术学院, 四川 广汉 618307)

摘要: 为解决快速存取记录器(quick access recorder, QAR)数据包含大量噪声影响飞行数据分析的问题, 提出一种基于长短期记忆网络(long short-term memory, LSTM)和卡尔曼滤波(Kalman filtering, KF)的组合降噪方法。运用拉依达准则对数据进行预处理, 基于 LSTM 建立模型中的状态方程, 结合 KF 对 QAR 数据进行实时在线估计, 并采用国产 ARJ21 飞机飞行数据进行仿真实验。结果表明: 该方法对实时数据的适应性优于单纯采用 LSTM 方法, 对动力学模型的依赖小于传统滤波法, 对 QAR 数据降噪处理的精度更高、降噪效果更好。

关键词: 长短期记忆网络; KF; QAR 数据; 数据降噪; ARJ21 飞机

中图分类号: TP391.9 **文献标志码:** A

Research on Data Denoising of Fast Access Recorder Based on Long-short Term Memory Network and Kalman Filter

Yang Junli, Yang Peida, Qian Yu, Xiao Yi

(School of Flight Technology, Civil Aviation Flight University of China, Guanghan 618307, China)

Abstract: In order to solve the problem that the quick access recorder (QAR) data contains a lot of noise, which affects the flight data analysis, a network based on long short-term memory (LSHM) and Kalman filtering (KF) is proposed. In this paper, the data is preprocessed by using the Leida criterion, and the state equation of the model is established based on LSTM, and the real-time online estimation of QAR data is carried out by combining with the Kalman filter, and the simulation experiment is carried out by using the flight data of the domestic ARJ21 aircraft. The results show that the adaptability of the proposed method to real-time data is better than that of the LSTM method, the dependence on the dynamic model is less than that of the traditional filtering method, and the proposed method has higher accuracy and better noise reduction effect for QAR data.

Keywords: LSTM model; Kalman filtering; QAR data; data noise reduction; ARJ21 aircraft

0 引言

国产 ARJ21 飞机和 C919 飞机的运营逐渐步入正轨, 标志着我国民用航空器的设计及运行能力不断提升。快速存取记录器(QAR)可以连续记录飞机数百小时的飞行数据, 完整反映了飞机系统的实时状态, 对飞行安全起着至关重要的作用^[1]。

由于飞机传感器的工作状态会受到外界环境和数据记录系统自身的精度、可靠性等因素的影响, QAR 数据存在着噪声、数据异常和数据缺失等问题, 对后续的数据处理和研究带来了很大困扰; 因此, 提高 QAR 数据的可靠性显得越发重要^[2-4]。

现有 QAR 数据处理主要集中在基于滤波的数据处理和基于时间序列的数据预测方面。欧阳舒歆等^[5]采用时间序列对 QAR 数据特征进行预测, 提出了一种基于卷积 VAE 和多头自注意力-LSTM 模型, 准确快速地检测到异常数据。钱宇等^[6]提出一种自

适应无迹卡尔曼滤波, 以无迹卡尔曼滤波为基础, 结合 Sage-Husa 噪声估计器对系统噪声进行实时预测和修正, 有效地解决了系统噪声时变问题。Deo 等^[7]采用无迹卡尔曼滤波对 QAR 参数降噪优化, 使其更好作用于飞行性能监控。张鹏等^[8]采用长短期记忆网络(LSTM)与注意力机制融合, 提出一种融合卷积神经网络与 LSTM 的双通道融合模型, 更有效地提取数据特征。张鹏等^[9]提出一种改进蝴蝶优化算法, 能够对飞行姿态角及角速度的 QAR 数据进行修正。

综上, 传统的卡尔曼滤波法(KF)对非线性系统的处理效果不够好, 改进的扩展卡尔曼滤波的系统稳定性较差, 时间序列法完全依赖历史数据, 缺乏对实时数据的适应性; 因此, 笔者提出一种基于神经网络和 KF 的组合模型估计方法对国产民机 QAR 数据进行降噪处理, 利用神经网络学习历史数据建

收稿日期: 2024-11-11; 修回日期: 2024-12-10

基金项目: 国家自然科学基金民航联合基金重点项目(U2133209); 民航安全能力建设基金(2022-239); 中央高校基本科研业务费专项资金资助(TD2025CZ02); 大学生创新创业训练计划项目(S202410624128)

第一作者: 杨军利(1972—), 男, 陕西人, 硕士。

立状态方程，再结合卡尔曼滤波进行状态的实时更新，是一种融合了机器学习和滤波方法的 QAR 数据降噪方法。

1 飞行数据预处理

在数据记录过程中，常有设备因自身或是外部的干扰导致数据出现空缺值、离群值^[10]。数据预处理的目的是为了处理这些异常值，提高数据的可靠性。拉依达准则是数据处理最常用的方法之一，常用来检测数据的粗大误差^[11]。粗大误差会直接影响观测的结果，它的产生是随机且难以避免的^[12]。

拉依达准则又称 3σ 准则，先假设样本数据中只含有随机误差，计算出样本数据的均值 μ 和标准偏差 σ 。按照一定的概率确定一个区间，只要超过这个区间的误差就可以认为是粗大误差，区间通常取 $\mu \pm 3\sigma$ ，所对应的置信概率为 99.7%，也就是数据不在区间 $(\mu - 3\sigma, \mu + 3\sigma)$ 内的为粗大误差，超出这个范围的概率极小，仅有 0.3%。超出范围的数据即为异常数据，应予剔除。假设一组正态分布数据为 (x_1, x_2, \dots, x_n) ，均值为：

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

标准差 σ 为：

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

第 i 个数据与均值之差 v_i 为：

$$v_i = x_i - \bar{x} \quad (3)$$

当某数据与均值之差 v_i 超出置信区间，满足式(4)时，可认为该数据为异常数据。

$$v_i > 3\sigma, \quad v_i < -3\sigma \quad (4)$$

2 数学模型

2.1 LSTM 模型

LSTM 是循环神经网络 (recurrent neural network, RNN) 的升级，可以解决 RNN 存在的长期依赖问题。LSTM 是一种时间循环神经网络，对长时间序列学习过程中因矩阵多次相乘而出现的梯度消失和梯度爆炸问题处理效果更好^[12]。LSTM 由遗忘门、输入门和输出门构成了门控系统，通过这些“门”的相互协调配合控制着信息流入和流出细胞状态的能力。LSTM 神经网络的结构如图 1 所示。

在遗忘层中， h_{t-1} 是系统在上一时刻的输出数据， x_t 是系统当前时刻的输入数据， h_t 是系统当前

时刻的输出数据， f_t 是系统在当前时刻的遗忘门，它决定了系统上一层的输出被传递到当前层的概率。遗忘门状态更新公式为：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

式中： σ 为 Sigmoid 函数； W_f 为遗忘门的权重值； b_f 为遗忘门的偏差量。

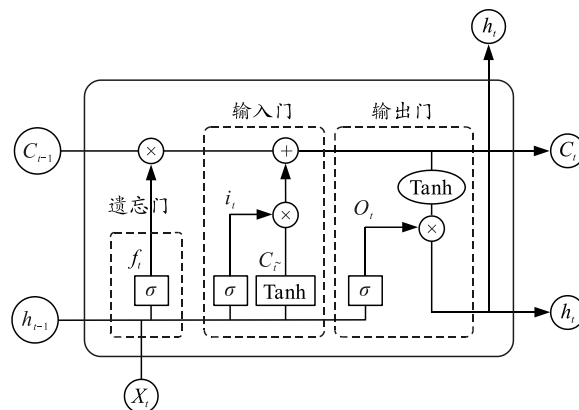


图 1 LSTM 神经网络结构

在输入层中， C_{t-1} 是系统在上一时刻的状态信息， C_t 是系统在当前时刻的状态信息， i_t 是系统在当前时刻的输入门，它决定了当前时刻的输入 x_t 有多少保存到系统当前时刻的状态信息 C_t 中。系统先通过 Sigmoid 函数决定哪些信息需要更新，再通过 Tanh 函数来添加新的输入信息。输入门状态更新为：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i); \quad (6)$$

$$C_t = \text{Tanh}(W_C \cdot [h_{t-1}, x_t] + b_C); \quad (7)$$

$$C_t = f_t * C_{t-1} + i_t * C_t \quad (8)$$

式中： W_i 为输入门的权重值； b_i 为输入门的偏差；Tanh 为双曲正切函数。

在输出层中，通过 Sigmoid 函数把 h_{t-1} 和 x_t 的线性组合结果转换成输出 o_t ；再通过 Tanh 函数将系统计算得到的新状态信息 C_t 与输出结果 o_t 相乘，得到当前时刻的输出值 h_t 。输出门状态更新为：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o); \quad (9)$$

$$h_t = o_t * \text{Tanh}(C_t) \quad (10)$$

式中： W_o 为输出门的权重值； b_o 为输出门的偏差。

2.2 KF 模型

在测量数据时会随机地产生误差，被称为噪声。噪声会影响数据真值，轻度影响会让信号变形、失真，重度影响会导致数据不可用，滤波的作用就是减小噪声对数据真值的影响^[13]。

KF 可依据上一时刻的状态估计值和当前时刻的观测值,对当前时刻的状态值进行最优估计。KF 对平稳随机过程和非平稳随机过程都有很好的降噪效果,被广泛应用在各个领域^[14]。一个离散的线性状态方程和观测方程为:

$$x_k = Ax_{k-1} + Bu_k + w_k; \quad (11)$$

$$z_k = Hx_k + v_k. \quad (12)$$

式中: x_k 为当前时刻系统的状态变量; x_{k-1} 为上一时刻系统的状态变量; z_k 为当前时刻系统的观测值; u_k 为当前时刻系统的控制输入; A 为系统的状态转移矩阵; B 为系统的控制输入矩阵; H 为系统的量测矩阵; w_k 为系统的过程噪声,系统从上一时刻的状态进入当前时刻的状态时,会存在外界因素的干扰; v_k 为系统的量测噪声,任意测量仪器都会存在一定的误差,且过程噪声 w_k 和量测噪声 v_k 是相互独立的高斯白噪声。

系统在 k 时刻的状态是由 $k-1$ 时刻的系统状态和 k 时刻的控制输入预测得到的,即 x_k 是由 x_{k-1} 和 u_k 预测得到的。数学公式如下:

$$x_k = Ax_{k-1} + Bu_k. \quad (13)$$

式中 x_k 为系统在当前时刻的预测状态,其协方差矩阵用 P_k 表示:

$$P_k = AP_{k-1}A^T + Q. \quad (14)$$

式中: P_{k-1} 为 x_{k-1} 的协方差矩阵; A^T 为 A 的转置矩阵; Q 为系统过程噪声的协方差矩阵。

结合 k 时刻系统状态的预测值和 k 时刻的观测值,可以得到 k 时刻系统的最优状态,即:

$$x_k = x_k + K_k(z_k - Hx_k). \quad (15)$$

式中: x_k 为系统在 k 时刻的最优估计状态; K_k 为系统在 k 时刻的卡尔曼增益:

$$K_k = \frac{H^T P_k}{HP_k H^T + R}. \quad (16)$$

更新 x_k 的协方差矩阵 P_k 为:

$$P_k = (I - HK_k)P_k. \quad (17)$$

式中 I 为单位矩阵。

2.3 基于 LSTM 和 KF 的混合模型

笔者利用 LSTM 基于历史数据学习到完整的状态方程,再应用 KF 结合实时的观测得到最优估计状态。用事先训练好的 LSTM 作为 KF 的预测方程,随后的更新步骤与 2.2 节相关内容相同,该组合模型的降噪流程如图 2 所示。

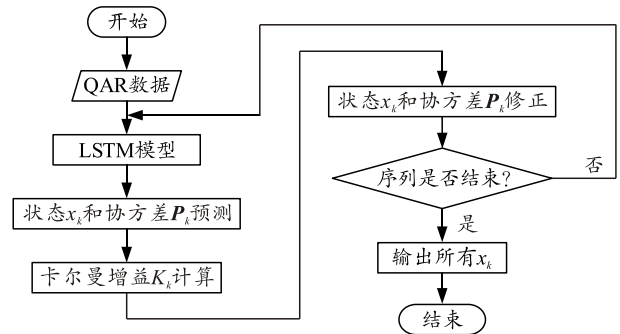


图 2 LSTM-KF 组合模型降噪流程

3 实验结果与分析

采用国产 ARJ21 飞机的实际运行数据进行仿真实验,共有 20 组 QAR 数据,每组 QAR 数据都记录了 13 个飞行阶段和包括马赫数、海拔高度、经纬度、时间信息等在内的各种飞行参数。选取俯仰角和滚转角这 2 个飞行参数验证模型的可行性。

3.1 样本数据检测与剔除

QAR 数据中存在很多噪声、异常和缺失等质量问题,给后续的数据挖掘、分析带来了很多问题;因此,非常有必要在降噪前对 QAR 数据的粗大误差进行处理。

在使用拉依达准则剔除粗大误差数据之前要检验样本数据是否满足拉依达准则的使用条件,要求数据符合正态分布且数据量足够大。QAR 数据样本数据量足够大,满足数据要求。图 3 为参数风向未经拉依达准则剔除粗大误差和剔除粗大误差后数据对比图。

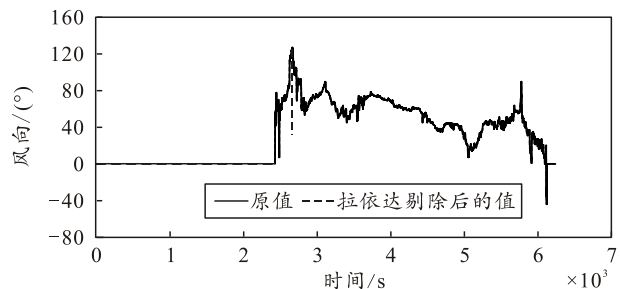


图 3 拉依达准则剔除粗大数据

从图 3 中可以看出:剔除粗大误差后的数据更简单,更稳定。

3.2 LSTM-KF 与 LSTM 降噪对比

选取某航空公司 ARJ21 飞机的 20 组 QAR 数据作为数据样本, QAR 数据按照飞行阶段参数 FLIGHT_PHASE 可以划分为飞机的爬升、巡航和下降 3 个阶段。选取常用的飞机性能监控参数俯仰角和滚转角的数据作为本实验的数据库^[7],将样本数

据的前 80%作为训练集，后 20%作为测试集，训练数据经预处理后进行实验验证。

将本文中实验方法与完全依赖历史数据的时序神经网络 LSTM 估计方法进行对比，分别得到了俯仰角和滚转角这 2 个参数在爬升、巡航和下降 3 种情况时的估计数据。实验结果证明了运用 LSTM-KF 模型对 QAR 数据进行降噪处理，更能提高数据的可靠性。

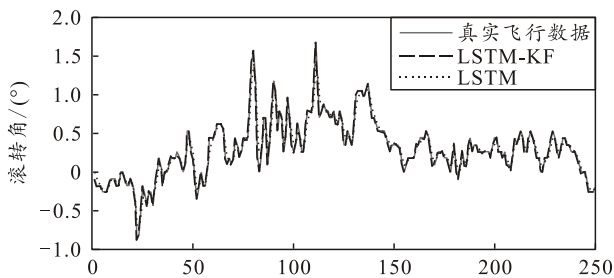
用平均绝对误差 (mean absolute error, MAE)、均方根误差 (root mean square error, RMSE) 对 LSTM-KF 和 LSTM 的估计效果进行评估, MAE 和 RMSE 的值越小, 说明模型估计的效果越好。其评价指标为:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|; \quad (18)$$

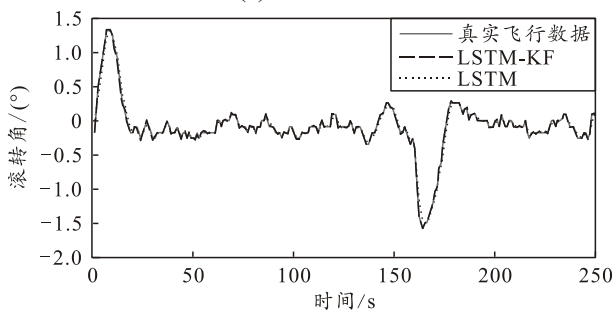
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}。 \quad (19)$$

LSTM-KF 较 LSTM 的优势在于加入了对历史数据的实时更新, 使实验结果更接近于真实数据。根据不同飞行数据的真实情况, 2 组数据的系统过程噪声和量测噪声的协方差矩阵 Q 和 R 已设置为适合参数的值。分别选取滚转角、俯仰角各 250 个数据, 验证算法的可行性。将数据分别输入 2 个模型, 经过 100 次迭代, 得到不同飞行阶段不同参数的估计结果。

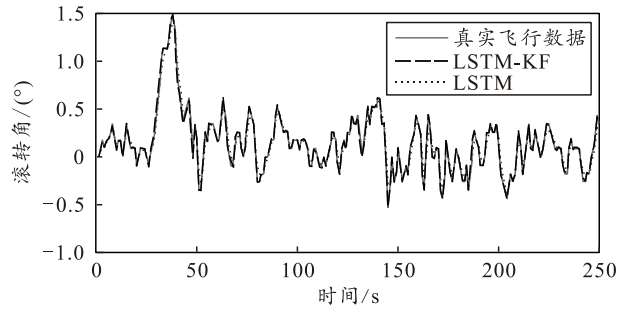
爬升、巡航和下降阶段参数滚转角 (roll angle, ROLL) 的估计结果如图 4 和 5 所示。



(a) 爬升阶段

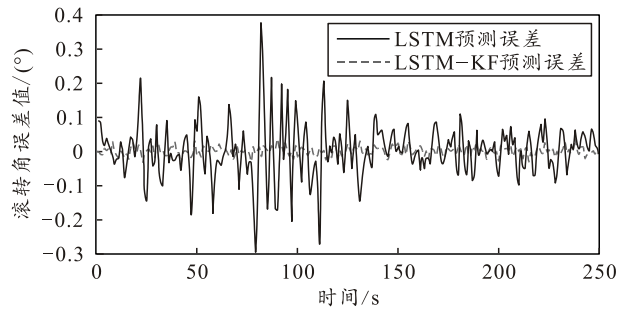


(b) 巡航阶段

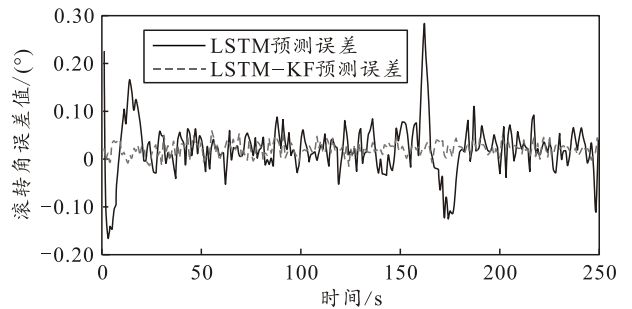


(c) 下降阶段

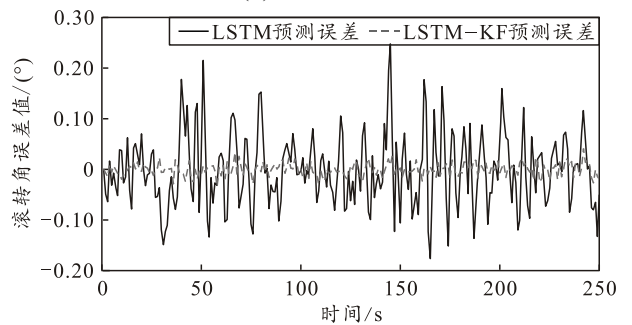
图 4 参数 ROLL 在不同飞行阶段的降噪效果



(a) 爬升阶段



(b) 巡航阶段



(c) 下降阶段

图 5 参数 ROLL 在不同飞行阶段的降噪误差对比

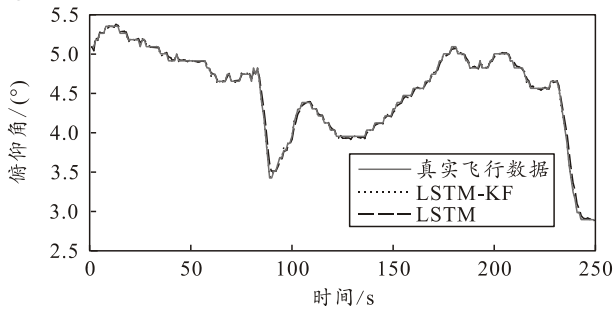
从图 4 和 5 可以看出: 参数 ROLL 经过 LSTM-KF 组合模型处理后的数据曲线与单一 LSTM 模型相比更平滑, LSTM 模型在 3 个阶段的预测误差值均大于 LSTM-KF 组合模型。在爬升阶段和下降阶段, ROLL 数据迅速变化, LSTM 模型的预测误差变化明显大于 LSTM-KF 组合模型, 说明 LSTM 模型的预测结果不如 LSTM-KF 组合模型稳定。MAE 和 RMSE 对比如表 1 所示。

表 1 参数滚转角降噪效果对比

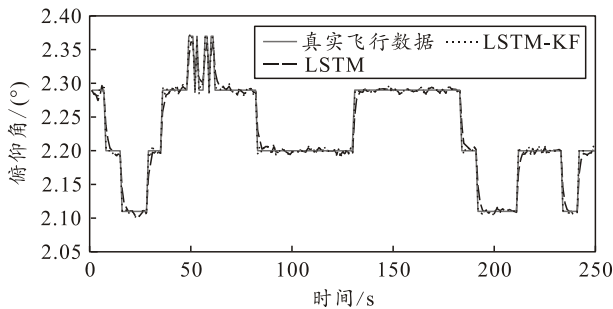
评价指标	爬升		巡航		下降	
	LSTM-KF	LSTM	LSTM-KF	LSTM	LSTM-KF	LSTM
MAE($\times 10^{-2}$)	1.142	5.495	1.080	3.452	1.018	5.655
RMSE($\times 10^{-2}$)	1.424	7.842	1.362	5.135	1.262	7.214

从表 1 中可以看出, LSTM-KF 组合模型的平均绝对误差和均方根误差的数值均小于 LSTM 模型, 说明 LSTM-KF 组合模型更贴近于真实数据。

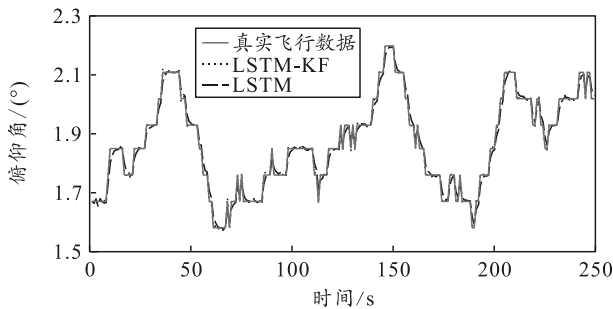
在爬升、巡航和下降阶段得到俯仰角 (pitch angle, PITCH) 参数的估计结果如图 6 和 7 所示。



(a) 爬升阶段

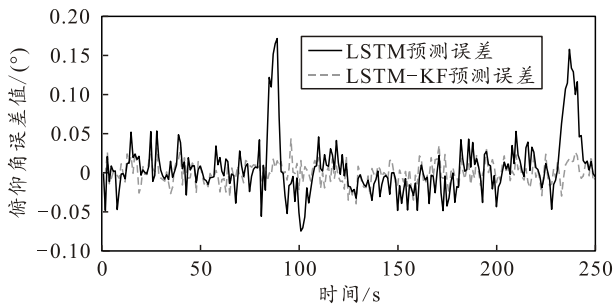


(b) 巡航阶段

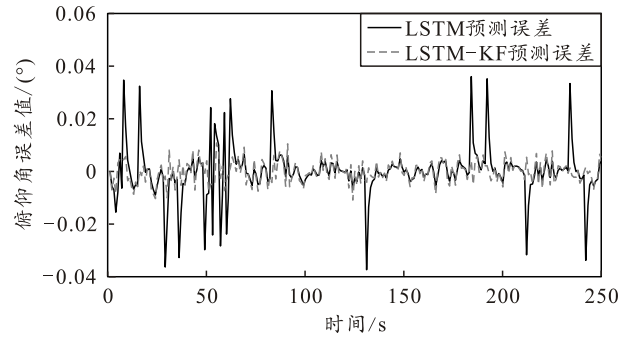


(c) 下降阶段

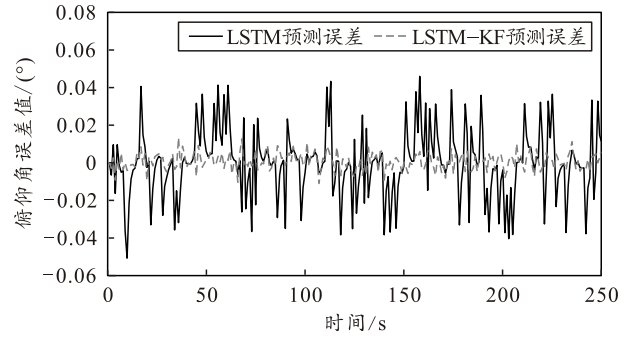
图 6 参数 PITCH 在不同飞行阶段的降噪效果



(a) 爬升阶段



(b) 巡航阶段



(c) 下降阶段

图 7 参数 PITCH 在不同飞行阶段的降噪误差对比

从图 6 和 7 可以看出: 参数 PITCH 经 LSTM-KF 组合模型处理过的数据与 LSTM 模型相比更接近原始数据, 且 LSTM-KF 组合模型每个飞行阶段的平均绝对误差和均方根误差均小于 LSTM 模型。说明 LSTM-KF 组合模型对原始数据的处理更好的抑制了噪声, 保留了原始数据, 提高了数据的可靠性。

表 2 详细列出 LSTM-KF 组合模型和 LSTM 模型在爬升、巡航、下降数据集上的整体性能。

表 2 参数 PITCH 降噪效果对比

评价指标	爬升		巡航		下降	
	LSTM-KF	LSTM	LSTM-KF	LSTM	LSTM-KF	LSTM
MAE($\times 10^{-3}$)	14.286	37.989	4.213	9.739	4.504	18.211
RMSE($\times 10^{-3}$)	11.344	23.607	3.407	5.489	3.611	12.934

结果表明: LSTM-KF 组合模型在 3 个数据集上的 MAE、RMSE 都优于 LSTM 模型, 且 LSTM-KF 组合模型性能在 QAR 数据不同阶段都有很好的效果。

4 结论

为更好地解决 QAR 数据的降噪问题, 笔者将传统的滤波方法和神经网络相结合, 提出一种 LSTM 和 KF 的组合模型对飞行参数进行状态估计的方法。选用国产 ARJ21 飞机的飞行数据进行仿真实验, 所得结论: