利用概率神经网络推断海洋输运统计量

Martin T. Brolly

爱丁堡大学数学学院与麦克斯韦数学科学研究所，King's Buildings，爱丁堡 EH9 3FD，英国

2023年6月5日

arXiv:2303.11480v2 [physics.ao-ph] 2 Jun 2023

摘要

我们利用概率神经网络和来自全球漂流浮标计划的拉格朗日观测，对海洋表层漂流浮标的单粒子转移概率密度函数（pdf）进行建模。转移概率密度由高斯混合表示，其参数（权重、均值和协方差）是纬度和经度的连续函数，并通过最大化观测到的漂流浮标轨迹的似然确定。该方法全面描述漂流浮标动力学，使我们能够模拟漂流浮标轨迹，并估计大量动力学统计量，而无需重新查阅原始数据。作为示例，我们计算了四天内平均位移和横向扩散率的全球估计。我们使用概率评分规则，将模型与常用的转移矩阵模型进行比较。我们的模型在全球范围以及三个特定区域均优于其他模型。利用我们的模型模拟漂流浮标释放实验，结果显示副热带环流中形成了密集簇，这与先前关于垃圾带形成的研究一致。神经网络模型的优势在于，它能够提供连续空间表示，无需对空间进行离散化，从而克服处理非均匀数据所面临的挑战。我们采用的数据驱动概率建模方法适用于流体动力学和海洋学中的许多其他问题。

1 引言

湍流流体的运动可通过扩散、能谱和速度结构函数等动力学统计量来有效表征（例如，Batchelor 1953；Monin & Yaglom 1971）。在海洋学中，人们投入了大量精力，试图根据观测值推断这类统计量（例如，LaCasce 2008；van Sebille 等 2018）。在许多情况下，这些推断任务可以转化为条件概率密度估计问题。例如，估计单粒子扩散与估计条件密度有关

$$p\left(\boldsymbol{X}(t+\tau)-\boldsymbol{X}(t) \mid \boldsymbol{X}(t), t, \tau\right), \tag{1}$$

其中，$\boldsymbol{X}(t)$ 表示粒子在 t 时刻的位置，扩散是该分布的方差。类似地，速度结构函数是条件

密度

$$p\left(\boldsymbol{u}(\boldsymbol{x}_1,\, t) - \boldsymbol{u}(\boldsymbol{x}_2,\, t) \mid \boldsymbol{x}_1,\, \boldsymbol{x}_2,\, t\right), \tag{2}$$

其中，$\boldsymbol{u}(\boldsymbol{x},\, t)$ 是位置 $\boldsymbol{x}$ 处的流体速度。通过估计如 (1) 和 (2) 所示的完整条件密度，可以同时估计多个相关统计量。例如，(1) 完整描述了单粒子位移统计量，而 (2) 则编码了所有阶数的速度结构函数，从而提供两点欧拉速度统计。由此可见，准确估计这些条件密度并非易事。

本文探讨一种用于条件密度估计的专门工具——混合密度网络（MDN）（Bishop 1994），并测试其从观测值中学习流体统计量的表现。混合密度网络是一类机器学习模型，它将人工神经网络与概率混合模型相结合，用于表示条件密度（Bishop 2006）。近年来，这类模型的应用迅速增加，涵盖多个领域中的各种降阶建模和仿真任务，包括流体流动的代理建模（Maulik et al. 2020）、海洋模型中亚网格动量强迫的参数化（Guillaumin & Zanna 2021）、流行病学中复杂随机模型的仿真（Davis et al. 2020）、化学反应网络的多尺度模型（Bortolussi & Palmieri 2018），以及湍流燃烧大涡模拟中的亚网格尺度闭合（Shin et al. 2021）。

我们重点利用全球漂流浮标计划（Lumpkin & Centurioni 2019）收集的拉格朗日轨迹数据，学习海洋近表层的单粒子转移密度（1）。转移密度模型能够为海洋中的每个位置提供漂流浮标从该位置出发后位移的概率预测。我们表明，混合密度网络模型优于基于乌拉姆方法（Ulam 1960，Froyland 2001）的现有漂流浮标动力学随机模型，也优于另一种简单的基准模型；此外，该模型无需为这类模型设计合适的空间离散化，从而消除了这方面的困难。

根据转移密度，可以推导出一系列单粒子统计量的估计值。例如，我们给出了以初始位置 $\boldsymbol{X}_0$ 为变量的四天平均位移分布图，以及横向扩散率分布图。在某些区域，转移密度会产生高度非高斯的统计量。通过计算完整模型与简化高斯模型之间的 Kullback–Leibler 散度，我们量化并绘制了漂流浮标位移的非高斯性分布图。

MDN模型还为漂流浮标动力学的离散时间马尔可夫过程模型提供了基础，为众多研究中采用的马尔可夫链模型（Maximenko等人 2012年，van Sebille 等 2012年，Miron 等人 2017、2021）提供了一种连续空间替代方案。我们以均匀网格上的初始位置为起点，对漂流浮标进行为期十年的全球模拟，并再现了以往研究中发现的副热带环流中的“垃圾带”。

本文结构如下。第 2 节讨论条件密度估计和条件统计量估计。第 3 节介绍混合密度网络。第 4 节描述基于漂流浮标观测的单粒子转移密度 MDN 模型。我们将其性能与替代模型进行比较，展示推导出的单粒子统计量，并模拟副热带环流中漂流浮标的聚集。第 5 节总结全文，并提出混合密度网络可能有用的其他问题。

2 条件建模

回归旨在对 $\mathbb{E}[\boldsymbol{Y} \mid \boldsymbol{X}]$ 建模，其中 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是随机变量；条件建模（或条件密度估计，CDE）则是推断完整条件概率密度 $p(\boldsymbol{Y} \mid \boldsymbol{X})$ [1] 的任务。通过对条件密度而非仅仅条件均值进行建模，我们纳入了关于 $\boldsymbol{Y} \mid \boldsymbol{X}$ 变异性的信息；不仅如此，条件模型还能够捕捉偏度、超额峰度和多峰性。在单点估计因固有变异性而不足以满足需求，且人们关注非高斯统计量（包括与罕见事件相关的统计量）的应用中，这种对条件统计量的全面描述十分有价值。条件模型可有两种用途：(i) 用作随机代理模型（或模拟器）；(ii) 作为估计条件统计量的工具。

参数化条件模型（例如混合密度网络）假设，对于 $\boldsymbol{X}$ 的每个可能取值，$\boldsymbol{Y} \mid \boldsymbol{X}$ 的分布都属于某个特定的参数分布族，即

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \rho(\boldsymbol{Y}\,;\, \boldsymbol{\theta}(\boldsymbol{X})), \tag{3}$$

其中，$\rho(\,\cdot\,;\,\boldsymbol{\theta})$ 是由 $\boldsymbol{\theta}$ 参数化的一族分布所对应的概率密度。在这种情况下，不仅必须选择 ρ 的形式，还必须通过某种 $\boldsymbol{\theta}(\boldsymbol{X})$ 表示来建模其对条件变量的依赖关系。

2.1 条件统计量估计

给定数据 $\{\boldsymbol{X}_i,\ \boldsymbol{Y}_i\}$，估计条件统计量 $\mathbb{E}[\boldsymbol{f}(\boldsymbol{Y}) \mid \boldsymbol{X}]$ 的一种标准方法是$\widehat{}$先对 $\boldsymbol{X}$ 进行离散化（或“分箱”），并针对离散化变量 $\boldsymbol{X}$ 的每个取值生成局部估计 $\mathbb{E}[\boldsymbol{f}(\boldsymbol{Y})](\boldsymbol{X})$，通常采用蒙特卡罗估计，使得

$$\widehat{\mathbb{E}[\boldsymbol{f}(\boldsymbol{Y})]}(\tilde{\boldsymbol{X}}) := \frac{\sum_i \boldsymbol{f}(\boldsymbol{Y}_i)\, \mathbb{1}_B(\boldsymbol{X}_i)}{\sum_i \mathbb{1}_B(\boldsymbol{X}_i)}, \tag{4}$$

其中，$\mathbb{1}_B$ 是 B 的示性函数；B 是 $\boldsymbol{X}$ 中离散化后的取值为 $\tilde{}\boldsymbol{X}$ 的那些值所构成的集合。要使估计式 (4) 有用，必须对 $\boldsymbol{X}$ 的定义域进行适当的离散化设计，在离散化足够精细、能够解析 $\boldsymbol{X}$ 中的细节，与分箱足够大、能包含足够多的数据以使这些估计的方差合理偏小之间取得平衡。当数据稀疏或数据密度高度不均匀时，这尤其具有挑战性。

条件建模提供了一种替代方法：首先构建一个如式 (3) 所示、关于 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$均连续的条件密度模型，然后计算估计值

$$\mathbb{E}_{\mathcal{M}}[\boldsymbol{f}(\boldsymbol{Y}) \mid \boldsymbol{X}] := \int \boldsymbol{f}(\boldsymbol{Y})\, \rho(\boldsymbol{Y}\,;\, \boldsymbol{\theta}(\boldsymbol{X}))\, \mathrm{d}\boldsymbol{Y} \tag{5}$$

对于定义域内任意的 $\boldsymbol{X}$ 值，都可以计算任意多个所需的统计量，而无需重新访问原始数据。在某些情况下，期望值 $\mathbb{E}_{\mathcal{M}}$ 可以通过闭式表达式计算。若不存在已知的此类表达式，则可以通过数值积分或蒙特卡罗方法计算期望值。由于这些计算只依赖于对建模得到的条件密度进行求值或从中抽样，因此不受数据稀疏性的限制。此外，对于给定的 $\boldsymbol{X}^*$，形式为 (4) 的估计仅由同一分箱中的观测值提供信息；而

[1] 我们仅考虑连续随机变量的情况。

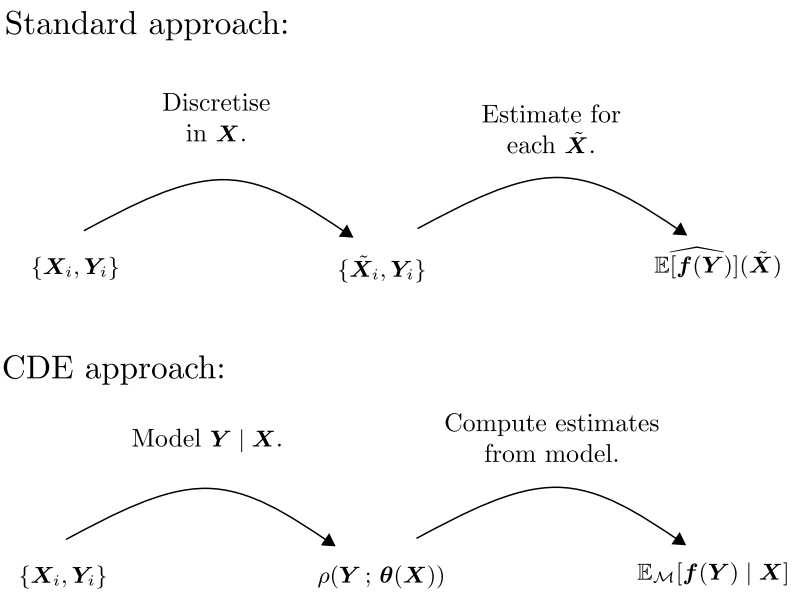


图 1：通过标准方法估计条件统计量，与先构建条件密度模型再进行估计的比较。

对于 $\boldsymbol{X}^*$，条件模型则使用所有观测值来拟合 $\rho(\boldsymbol{Y};\ \boldsymbol{\theta}(\boldsymbol{X}^*))$。图 1 中的示意图对比了标准方法与条件建模方法。

3 混合密度网络

混合密度网络（Bishop 1994, 2006）是一种条件模型，其中使用人工神经网络表示式 (3) 中的函数 $\boldsymbol{\theta}(\boldsymbol{X})$，而参数形式 $\rho(\cdot;\ \boldsymbol{\theta})$ 对应于混合分布。一般混合分布的密度为

$$\rho(\,\cdot\,;\,\boldsymbol{\theta}) = \sum_{i=1}^{N_c} \alpha_i\ \rho_i(\,\cdot\,;\,\boldsymbol{\theta}_i), \tag{6}$$

其中，N_c 是混合分布中的分量数，第 i^{th} 个分量的密度为 $\rho_i(\,\cdot\,;\ \boldsymbol{\theta}_i)$，其参数为 $\boldsymbol{\theta}_i$、$\boldsymbol{\theta} = [(\alpha_1,\ \ \boldsymbol{\theta}_1),\ \cdots,\ (\alpha_{N_c},\ \ \boldsymbol{\theta}_{N_c})]$，而 α_i 是受以下约束的分量权重

$$\sum_{i=1}^{N_c} \alpha_i = 1. \tag{7}$$

通常，分量密度 ρ_i 选自同一族，尤其是高斯分布，但也可以选择不同的分量。在高斯情形下，$\boldsymbol{\theta}_i$ 是条件均值和协方差。

$\boldsymbol{\theta}(\,\cdot\,)$ 的神经网络表示本身也是参数化的，其参数为 $\boldsymbol{w}$；因此，混合密度网络用 $\rho(\boldsymbol{Y};\ \boldsymbol{\theta}(\boldsymbol{X};\ \boldsymbol{w}))$ 对 $p(\boldsymbol{Y}\ \mid\ \boldsymbol{X})$ 进行建模。网络可以采用任意架构，

但常见的是采用具有非线性激活函数的多层感知机（Rumelhart 等，1985）（也称为全连接多层前馈神经网络）；在这种情况下，$\boldsymbol{w}$ 由权重和偏置组成。

对于条件模型，一种自然的损失函数可以量化模型对数据的拟合程度，即模型下观测值 $\mathrm{D} = \{\boldsymbol{X}_i, \boldsymbol{Y}_i\}$ 的负（条件）对数似然。在混合密度网络中，该函数为

$$\mathcal{L}(\boldsymbol{w}; \mathcal{D}) = \sum_i -\log \rho\left(\boldsymbol{Y}_i; \boldsymbol{\theta}(\boldsymbol{X}_i; \boldsymbol{w})\right). \tag{8}$$

训练　混合密度网络的训练，就是为神经网络的参数寻找最优值　s

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}; \mathcal{D}). \tag{9}$$

最小化负对数似然，等价于最大化训练数据的对数似然；在概率预测中，这也称为对数评分（Bernardo（1979）、Gneiting 与 Raftery，2007、Br¨ocker 与 Smith，2007）。此处的最大似然估计与更常见的情形有所不同：在后者中，给定观测数据 $\{\boldsymbol{Y}_i\}$，拟合的是 $p(\boldsymbol{Y})$ 的无条件模型——在这里，每个观测到的 $\boldsymbol{X}$ 值通常仅对应一个观测到的 $\boldsymbol{Y} \mid \boldsymbol{X}$ 值，而对于大多数 $\boldsymbol{X}$ 值，根本没有观测值。因此显然，对于每个 $\boldsymbol{X}$ 值，我们都不处于能够援引最大似然估计值渐近性质的大数据情形。参数化条件模型（3）的质量，关键取决于 $\boldsymbol{\theta}(\boldsymbol{X}; \boldsymbol{w}^*)$ 对 $\boldsymbol{Y} \mid \boldsymbol{X}$ 的分布如何随 $\boldsymbol{X}$ 变化的刻画是否准确。特别是，由于混合密度网络使用神经网络来建模 $\boldsymbol{\theta}(\boldsymbol{X})$，而神经网络具有很强的灵活性，因此若不采用正则化技术，混合密度网络通常会表现出较差的泛化能力。在下一节中，我们将采用一种广泛使用的正则化技术，称为提前停止（例如参见 Prechelt（2012））：将一小部分训练数据（称为测试集）排除在优化方案的迭代步骤之外，而用它来跟踪模型泛化误差估计值的变化（即在训练集之外的数据上计算的损失函数值）。这一方法基于如下经验启发：随着训练的进行，神经网络的泛化误差通常会先降至最低点，随后因过拟合而上升；提前停止策略则是在泛化误差被认为已达到最低点时终止模型训练。下一节将介绍我们实现该方法的具体细节。

4 海洋近表层单粒子统计的应用

本节介绍海洋表层漂流浮标（漂流浮标）的单粒子转移密度(1)混合密度网络模型。模型参数根据全球漂流浮标计划（Lumpkin & Centurioni 2019年）收集的轨迹数据推断得出。

4.1 数据

我们使用全球漂流浮标计划提供的质量控制的6小时插值数据集，其中包含位置（纬度和经度）以及海表温度。数据集还提供漂流浮标速度估计值，不过这些估计值是通过对位置

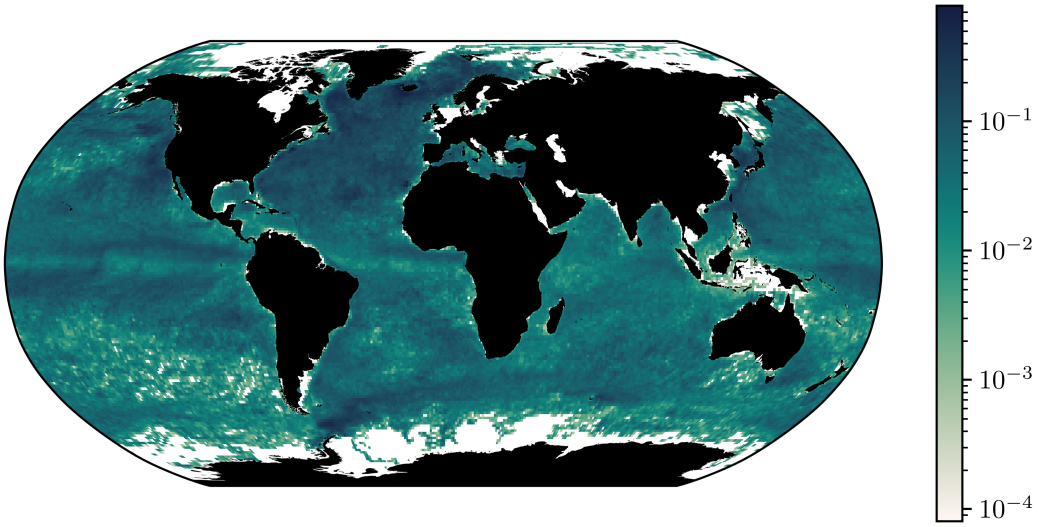


图2：每平方公里的漂流浮标观测数量。

测量数据进行简单的有限差分得到的。位置测量数据来自卫星定位，其时间间隔不均匀且存在误差。原始测量数据按照 Hansen & Poulain (1996) 的流程处理，包括剔除疑似异常值，并插值到规则的6小时区间。该插值方法是一种克里金法（Hansen & Herman 1989），假设数据受到不相关零均值噪声的污染，并对离散化位置过程的结构函数作出假设。我们将这一数据预处理可能存在争议之处作为结果的一个注意事项，并将插值数据视为真实值。在建模中，我们只使用位置观测。图2展示了每个 1° 纬度 $\times\, 1^\circ$ 经度方格中每平方公里记录的观测位移数量。这些数据记录于1989年至2021年，共包含23893条漂流浮标轨迹。我们将数据拆分为两部分：随机选取约一半（11946条）漂流浮标轨迹用于建模，其余数据留作验证。整个数据集包含超过1800万条6小时位移观测。

在第4.3节中，我们进行模型比较。我们针对具有全球覆盖的完整训练和验证数据集，以及三个受限区域，计算技巧评分。这三个区域在图3中标记为 A、B和 C，范围分别为20–50° W、30–50° N；145–175° E、20–40°N；以及110–130° W、 10° S–10° N。

4.2 模型

转移密度并未采用最一般的形式进行建模。相反，我们（i）首先考虑时间滞后 τ的一个固定值，因此转移密度可以写作

$$p(\boldsymbol{X}_{n+1} \mid \boldsymbol{X}_n), \tag{10}$$

其中 $\boldsymbol{X}_n = \boldsymbol{X}(t_0 + n\tau)$，并且 (ii) 假设过程 $\boldsymbol{X}(t)$ 是时间齐次的，因此 (1) 与初始时间 t 无关，(10) 与 n 无关且

$$p(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0) \tag{11}$$

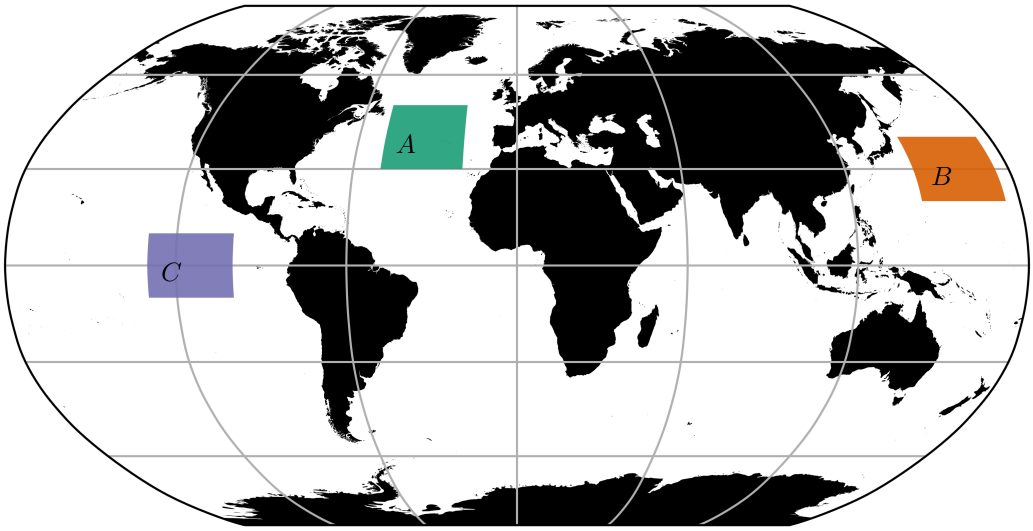


图3：第4.3节中用于模型比较的区域。

表示与 (10) 相同的信息，其中 $\Delta \boldsymbol{X}$ 是漂流浮标相对于上一个时间步所在位置的位移，记为 $\boldsymbol{X}_0$。通过假设时间齐次性，我们忽略海洋动力学中季节性和低频变化的影响。此外，若进一步假设马尔可夫性，则 (10) 足以构建一个描述漂流浮标位置的离散时间马尔可夫过程模型（$\boldsymbol{X}_n$）（Pavliotis 2014）。要使马尔可夫假设准确，必须恰当选择离散化时间尺度 τ。我们选择 4 天的时间尺度，这是因为此前根据漂流浮标数据估计，四个海洋盆地表层的拉格朗日速度去相关时间（或积分时间尺度）约为 2–3 天（Rupolo 2007）。

我们使用混合密度网络对式（11）进行建模——参见图4中的示意图。模型的输入为$\boldsymbol{X}_0$，以经纬度坐标表示；输出为一个高斯混合分布，其中包含 $N_c = 32$ 个混合分量，用于对 $\Delta \boldsymbol{X} \mid \boldsymbol{X}_0$ 进行建模，坐标同样以相对于 $\boldsymbol{X}_0$的经度和纬度（度）表示。因此，模型中的神经网络部分编码

$$\boldsymbol{\theta}(\cdot) = \left\{\alpha_i(\cdot),\ \boldsymbol{\mu}_i(\cdot),\ \mathbf{C}_i(\cdot)\right\}_{i=1}^{N_c} \tag{12}$$

使得

$$\begin{aligned} p(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0) = \sum_{i=1}^{N_c} \alpha_i(\boldsymbol{X}_0) \det\left(2\pi \mathbf{C}_i\left(\boldsymbol{X}_0\right)\right)^{-\frac{1}{2}} \\ \times \exp\left[-\frac{1}{2}\left(\Delta \boldsymbol{X} - \boldsymbol{\mu}_i\left(\boldsymbol{X}_0\right)\right)^{\mathrm{T}} \mathbf{C}_i^{-1}\left(\boldsymbol{X}_0\right)\left(\Delta \boldsymbol{X} - \boldsymbol{\mu}_i\left(\boldsymbol{X}_0\right)\right)\right], \end{aligned} \tag{13}$$

其中，$\boldsymbol{\mu}_i$ 和$\mathbf{C}_i$ 分别是混合成分 i的均值向量和协方差矩阵。混合成分的数量是一个可优化的超参数。我们选择$N_c = 32$，因为在使用混合密度网络进行的试验中发现，包含32个分量的混合分布具有足够的表达能力。

所选神经网络架构为标准多层感知机，包含六个隐藏（即内部）层。前四个隐藏层各有 256 个神经元，后两个各有 512 个。双曲正切激活函数 $\tanh(x)$ 应用于每个隐藏

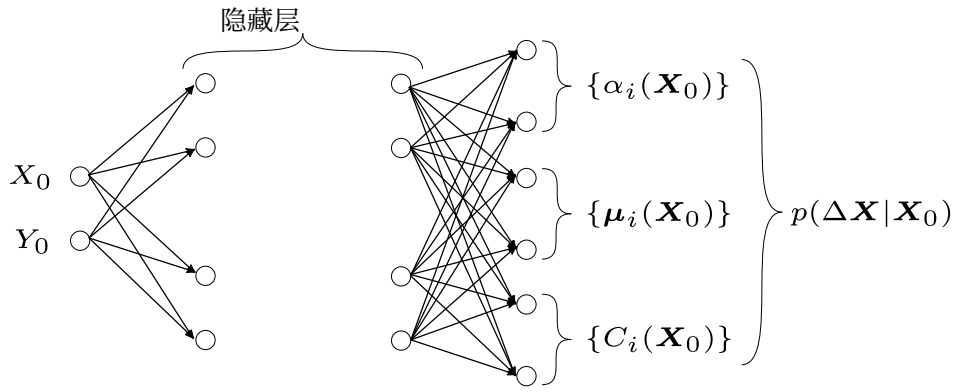


图4　：漂流浮标单粒子转移密度的混合密度网络模型示意图。

层。因此，隐藏层 i的活动为

$$\boldsymbol{h}_i = \tanh\left(W_i\boldsymbol{h}_{i-1} + \boldsymbol{b}_i\right) \tag{14}$$

对于 $i > 1$，以及

$$\boldsymbol{h}_1 = \tanh\left(W_1\boldsymbol{X}_0 + \boldsymbol{b}_1\right). \tag{15}$$

这里，$\boldsymbol{W}_i \in \mathbb{R}^{d_i\times d_{i-1}}$ 和 $\boldsymbol{b}_i \in \mathbb{R}^{d_i}$ 是对应于i^{th} 层的权重和偏置参数，该层有 d_i 个神经元。注意，$\boldsymbol{w} = \{\boldsymbol{W}_i,\ \boldsymbol{b}_i\}$。最后一层采用定制的激活函数，以确保 $\boldsymbol{\theta}$各分量满足自然约束。具体而言，对应于混合分量权重 $\boldsymbol{\alpha}$的神经网络输出采用 Softmax 激活函数 $a_{\text{sm}}(\boldsymbol{x}) = \exp(\boldsymbol{x}) / \sum_i \exp(x_i)$，从而确保这些权重为正并满足约束(7)。每个协方差矩阵$\mathbf{C}_i$ 均由一个下三角乔列斯基因子的各分量表示——通过取指数来确保对角元素为正。当 $N_c = 32$ 时，我们有 $\dim(\boldsymbol{\theta}) = 192$，神经网络参数（即权重和偏置）的总数为 $\dim(\boldsymbol{w}) = 690,880$。我们使用 Adam 算法（ Kingma & Ba 2015 ）最小化负对数似然损失函数(8)，以训练模型。需要指出的是，隐藏层的宽度数量也是超参数，我们在测试问题上经过实验后选定了这些超参数。本文不试图为这些参数寻找最优值。

与机器学习中的常见做法一样，我们在训练前对数据进行标准化（LeCun 等人，2012年）：分别对输入数据 $\{\boldsymbol{X}_{0i}\}$ 和输出数据 $\{\Delta\boldsymbol{X}_i\}$ 进行变换，即减去训练数据的均值，再除以训练数据中各分量的标准差，使变换后数据的每个分量均值为零、方差为一。虽然这种做法缺乏充分或令人满意的理论依据，但我们发现它确实显著提高了优化过程的数值稳定性。无论如何，我们采用的变换是可逆的，不过必须谨慎地对转移密度进行反向缩放，才能得到正确结果。例如，若将标准化后的变量记为 $\boldsymbol{X}_0$ 和 $\Delta\boldsymbol{X}$，则模型近似表示 $p(\Delta\boldsymbol{X} \mid \boldsymbol{X}_0)$，我们可以据此还原出单位正确的转移密度：

$$p(\Delta\boldsymbol{X} \mid \boldsymbol{X}_0) = \frac{p\left(\widetilde{\Delta\boldsymbol{X}} \mid \widetilde{\boldsymbol{X}_0}\right)}{\widehat{\text{std}}(\Delta X)\ \widehat{\text{std}}(\Delta Y)}, \tag{16}$$

其中，$\widehat{\text{std}}(\cdot)$ 表示训练数据的样本标准差。

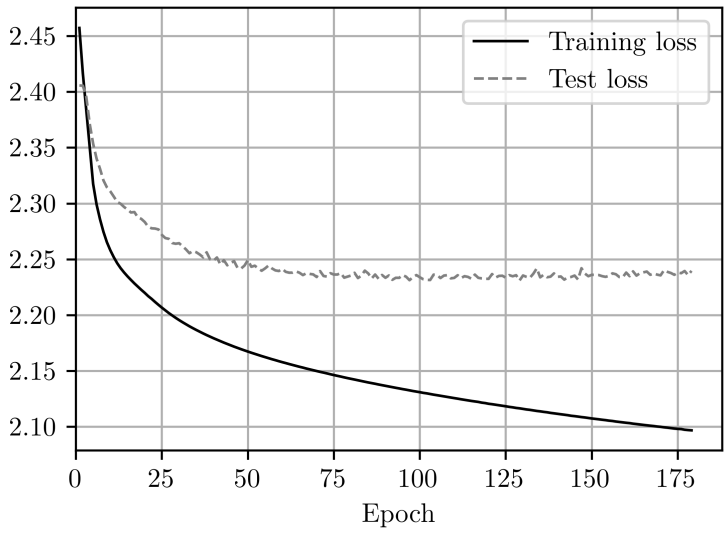


图5：优化过程中，MDN模型的训练损失和测试损失的变化。图中显示的损失为每个数据点的平均负对数似然（即式（8）的归一化形式），$\widetilde{\quad}$ $\widetilde{\quad}$以标准化变量 $\boldsymbol{X}_0$ 和 $\Delta\boldsymbol{X}$表示。

神经网络有一个与当前问题尤为相关的特点：它们难以表示周期函数（Liu 等人（2020））。由于我们使用经纬度坐标，转移密度模型应当在经度方向上具有周期性。然而，神经网络模型接收到的初始位置 $\boldsymbol{X}_0$ 只是 $\mathbb{R}^2$中的一个向量，因此其表示中并未内置球面域的概念。事实上，由于国际日期变更线两侧存在模型误差，MDN模型会在国际日期变更线处的 $p(\Delta\boldsymbol{X} \mid \boldsymbol{X}_0)$产生不连续性。为改善国际日期变更线处的连续性，我们采用了一种粗略的方法：将数据复制两次，一份向西平移 $360°$ 经度，另一份向东平移 $360°$。

该模型使用 Python、TensorFlow（TensorFlow Developers 2021）和 TensorFlow Probability（TensorFlow Probability Developers 2021）实现，并使用四块 NVIDIA Tesla V100 16GB GPU 并行训练。在用于构建模型的 50% 数据中，随机选取 90% 用于优化过程的迭代，10% 用于提前停止——我们分别将这两部分数据称为训练集和测试集。训练耗时约 90 分钟。图5展示了训练过程中训练集和测试集上的损失函数随训练轮次的变化。一个训练轮次指在 Adam 算法中将全部数据使用一次所需的迭代次数。优化所采用的停止标准（提前停止的一种示例）是：测试损失在此前 50 个训练轮次内没有下降。

4.3 模型评估与比较

由于MDN模型是概率模型，因此应使用概率预测的技巧评分来评估其性能，而不是采用确定性模型常用的性能指标，例如均方误差。如上所述，最小化负对数似然等同于最大化对数评分，因为对数评分正是对数似然。对数评分具有一些优良性质，即严格适当性（Br¨ocker & Smith 2007）和局部性（Du（2021））。事实上，对于连续变量而言，除仿射变换外，它是唯一一种光滑、局部且严格适当的评分规则（Bernardo（1979））。若某评分规则的期望

（相对于数据而言）仅在正确/完美模型（假设该模型存在）下达到最大值，则该评分规则是严格适当的。如果某评分规则仅取决于在观测数据处评估的预测概率分布值，而不取决于预测分布的其他特征（例如其形状），则该规则是局部的。为了进行验证，我们可以在验证数据集上计算对数评分。然而，对于离散/分类变量的预测，对数评分的值很容易解释；而对于连续变量，其值并没有直观意义，因为它指的是概率密度，其量纲与支撑集面积的倒数相同，这意味着对数评分的尺度取决于具体问题。另一方面，将对数评分作为模型间的相对评分时，其含义更容易理解——尤其是，模型间对数评分差值的均值反映了第一个模型相较于另一个模型，在观测结果上赋予的平均额外概率；这一差值以信息单位纳特（或在评分定义中使用 $\log_2$ 时的香农）衡量。对数评分之差在光滑变换下保持不变（Du（2021））；这尤其意味着，单位变换不会影响对数评分之差。因此，为了评估MDN模型，我们将其与替代模型进行比较。我们介绍两种替代模型：一种在文献中得到广泛应用，另一种则是本文提出的简单而合理的替代方案。我们还将其与简化版MDN模型进行比较，该模型仅包含一个混合成分，即 $N_c = 1$。所有模型的对数评分均在训练数据和验证数据上计算，以评估相对性能。

4.3.1 转移矩阵模型

先前的研究（Maximenko 等人，2012；van Sebille 等，2012；Miron 等人，2017、2021）使用乌拉姆方法（乌拉姆，1960；Froyland，2001），通过离散时间马尔可夫链对漂流浮标动力学进行建模。这需要将空间离散为分箱 $\{B_i\}$，并估计转移矩阵

$$P_{ij} = \mathbb{P}(\boldsymbol{X}_{n+1} \in B_j \mid \boldsymbol{X}_n \in B_i), \tag{17}$$

它是转移密度（10）的离散对应形式。事实上，马尔可夫链模型与我们的马尔可夫过程模型之间的主要区别在于，我们的模型在空间上是连续的。转移矩阵的元素通常通过图1所示的标准方法进行估计，其中我们有 $\boldsymbol{Y} = \boldsymbol{X}_{n+1}$ 和 $f(\boldsymbol{Y}) = \mathbb{1}_{B_j}(\boldsymbol{X}_{n+1})$——这对应于每个 P_{ij} 的最大似然估计，因此会使训练数据集上的对数评分最大化。请注意，可利用转移矩阵构造相应的转移密度，该密度在 $\boldsymbol{X}_n$ 和 $\boldsymbol{X}_{n+1}$ 中的网格单元上为分段常数，方法如下

$$p(\Delta\boldsymbol{X} \mid \boldsymbol{X}_0) = \frac{P_{ij}}{A(B_j)}, \quad \text{when } \boldsymbol{X}_0 \in B_i,\ \boldsymbol{X}_0 + \Delta\boldsymbol{X} \in B_j, \tag{18}$$

其中，$A(B_j)$是 B_j的面积[2]。这对于与空间连续模型进行比较至关重要。

马尔可夫链模型的一个优点是，对其长期行为的分析较为直接——可以研究转移矩阵的左右特征向量，以识别近似不变集，如 Miron 等人（2017）所做的那样。这种方法被称为特征向量方法（Froyland 等人，2014）。将这一分析扩展到连续空间情形

[2]为与MDN模型给出的转移密度保持一致，这些面积必须使用相同的变量计算，即经度度数乘以纬度度数。

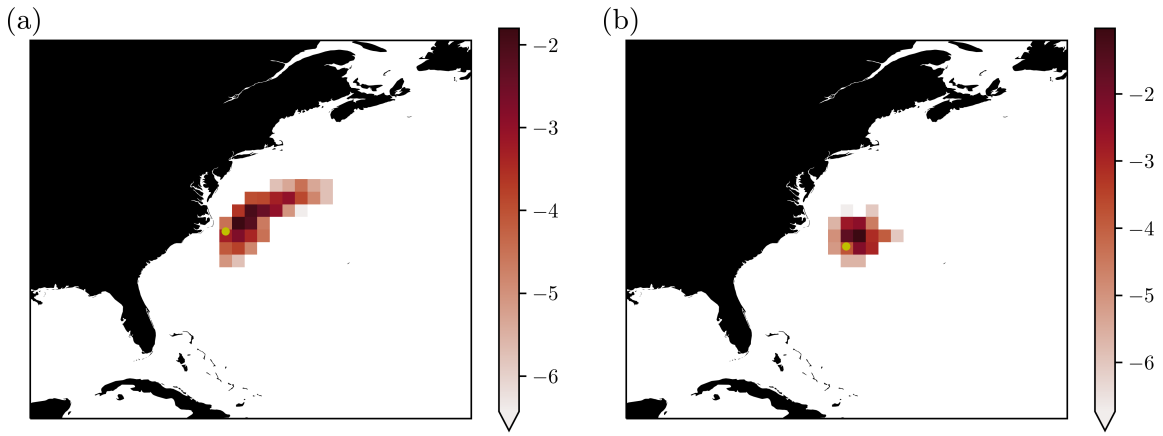


图 6：初始位置 $\boldsymbol{X}_0$ 对应的对数转移概率密度函数 $\log\, p(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0)$ 地图：(a) 位于墨西哥湾流内（34.85° N，74.50° W）；(b) 位于墨西哥湾流附近（33.67° N，72.55° W）。该结果由转移矩阵模型通过式（18）并使用 $\tau = 4$ 天得出。黄色圆点表示 $\boldsymbol{X}_0$。

使用我们的模型进行这一扩展（留待未来工作），需要计算相关佩龙–弗罗贝尼乌斯算子 P 的特征函数。该算子作用于概率密度函数，使其随时间向前演化，从而满足

$$p(\boldsymbol{X}_{n+1}) = \mathcal{P}(p(\boldsymbol{X}_n)) \tag{19}$$

$$:= \int_\Omega p(\boldsymbol{X}_n)\, p(\boldsymbol{X}_{n+1} \mid \boldsymbol{X}_n)\, \mathrm{d}\boldsymbol{X}_n. \tag{20}$$

另外值得注意的是，可以通过对转移密度进行数值积分来使用混合密度网络模型构建转移矩阵，也就是通过数值计算

$$P_{ij} = \int_{B_i} \int_{B_j} p(\boldsymbol{X}_{n+1} \mid \boldsymbol{X}_n)\, \mathrm{d}\boldsymbol{X}_{n+1}\, \mathrm{d}\boldsymbol{X}_n. \tag{21}$$

图 6 展示了转移矩阵模型通过式（18）得出的两种不同初始位置 $\boldsymbol{X}_0$ 对应的对数转移密度 $\log\, p(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0)$。第一个位置位于墨西哥湾流核心区域内的 34.85° N，74.50° W；第二个位置则紧邻墨西哥湾流之外，位于33.67° N，72.55° W。请注意，在这两种情况下，密度的支撑集都是训练数据中观测到发生转移的网格单元集合。换言之，在该模型下，转移到其他网格单元的概率为零。我们将在第 4.3.3 节回到这一点。

4.3.2 采用网格参数的高斯转移（GTGP）

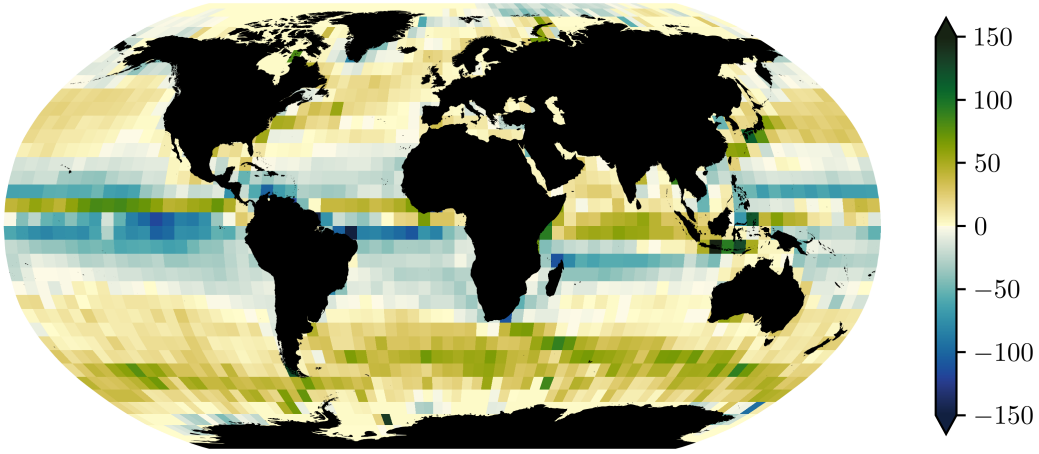
转移密度 (11) 的一个简单模型是：给定初始位置 $\boldsymbol{X}_0$，转移服从条件高斯分布，其条件均值和协方差由 $\boldsymbol{X}_0$ 的函数给出，并且这些函数在网格单元上是分段常数，即

$$\Delta \boldsymbol{X} \mid \boldsymbol{X}_0 \sim \mathcal{N}\left(\boldsymbol{\mu}\left(\boldsymbol{X}_0\right), \mathbf{C}\left(\boldsymbol{X}_0\right)\right), \tag{22}$$

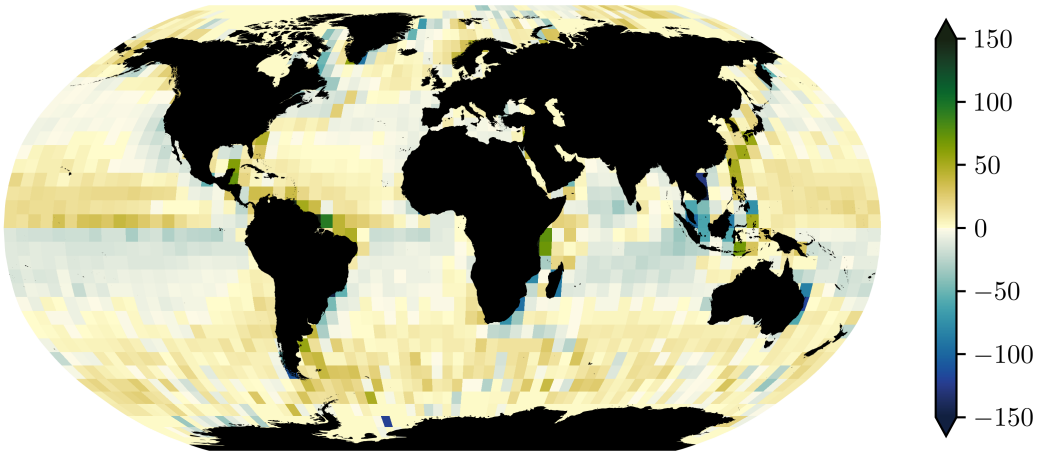
其中 $\boldsymbol{\mu}(\boldsymbol{X}_0)$ 和 $\mathbf{C}(\boldsymbol{X}_0)$ 在 $\boldsymbol{X}_0$ 中为分段常数。通过将观测值排序后分入各个分箱，并计算每个分箱的样本均值和样本协方差，来估计参数 $\boldsymbol{\mu}$ 和 $\mathbf{C}$。样本均值是 $\boldsymbol{\mu}$ 的最大似然估计，而样本

协方差与 $\mathbf{C}$ 的最大似然估计仅相差一个 $\frac{N-1}{N} \approx 1$ 的因子，其中 N 是给定分箱中的训练数据数量。因此，所采用的参数估计与使训练数据上的对数评分最大化的参数估计非常接近。

图7展示了具有规则 $1^\circ \times 1^\circ$ 经纬度网格、时间跨度为 $\tau = 4$ 天的GTGP模型所得位移均值随初始位置的变化。



(a) 4天纬向位移均值（km）。



(b) 4天经向位移均值（km）。

图7：时间跨度为 $\tau = 4$ 天的GTGP模型所得位移均值随初始位置的变化。

4.3.3 模型评分

我们计算了全球覆盖的完整训练集和验证数据集，以及区域 A、B 和 C 的技巧评分。对于转移矩阵模型和 GTGP 模型，都需要选择空间离散化方式；本文仅考虑正方形经纬度网格，

		TM	GTGP	MDN1	MDN32
全局：	训练	−1.61	−1.20	−1.07	−1.02
	验证	−1.74	−1.30	−1.14	−1.11
A:	训练	−1.78	−1.25	−1.35	−1.30
	验证	−1.89	−1.42	−1.40	−1.35
B:	训练	−1.95	−1.90	−1.91	−1.85
	验证	−2.16	−2.01	−2.01	−1.96
C:	训练	−1.93	−1.59	−1.63	−1.56
	验证	−2.00	−1.66	−1.61	−1.57

表1：各研究区域中转移矩阵和 GTGP 模型，以及单分量混合密度网络和完整的 32分量混合密度网络模型的训练与验证评分（参见图3中的地图）。评分为每个数据点的平均对数评分（即平均对数似然），根据变量 $\boldsymbol{X}_0$ 和 $\Delta\boldsymbol{X}$ 以其原始经度/纬度单位计算。

因此唯一需要选择的参数是网格单元边长。这一选择会影响模型的表现。采用相对较高分辨率的离散化时，模型在训练集上的得分相对较高，但泛化能力较差，这从验证数据上的低分可以看出。对于 GTGP 模型，当验证数据落在训练集中漂流浮标未曾经过的网格单元内时，就会出现问题，因为在没有数据的分箱中无法估算样本均值和协方差。一个简单的解决办法是，将未经过网格单元的 $\boldsymbol{\mu}$ 和 $\mathbf{C}$ 设为全局（或区域）估计值。按照上述方法估算转移矩阵时，转移矩阵模型的一个缺陷是，验证数据在模型下可能具有零概率，因此对数评分为负无穷。采用足够大的网格单元可以避免这种情况，但这会导致得分极低。另一方面，如果愿意直接舍弃在模型下概率为零的验证数据，那么使用更小的网格单元也能计算验证得分。这似乎过于宽松，因为随着网格单元大小趋近于零、越来越多的验证数据被忽略，转移矩阵模型的得分会不断提高。作为折中，我们将转移矩阵模型的网格单元大小固定为 $1^\circ \times 1^\circ$，即一些以往研究（van Sebille 等，2012）使用转移矩阵模型时采用的分辨率，并舍弃概率为零的验证数据——全球范围内舍弃的验证数据占 7%，区域 *A*、 *B* 和 *C* 分别占 2%、9% 和 4%。对于 GTGP 模型，我们使用网格搜索交叉验证来优化网格单元大小，以最大化验证得分：这是一种常见方法，即尝试模型超参数（此处为网格单元大小）的一系列取值，并选择能使验证得分达到最优的取值。找到的最优网格单元大小从区域 A 的 1.1° 到全球范围的 5° 不等。得分列于表 1。所有区域中，混合密度网络模型的表现都优于其他模型，其中 32分量模型的得分略高于单分量模型。需要注意的是，报告的得分恰好为负值——这并非惯例所致，而是因为对数概率密度通常为负值。得分越高越好。

4.4 结果

训练完成后，该模型至少可以有两种用途：(i) 推导单个漂流浮标位移统计量的估计值，以及 (ii) 模拟漂流浮标轨迹。不过，我们首先直接考察转移密度。

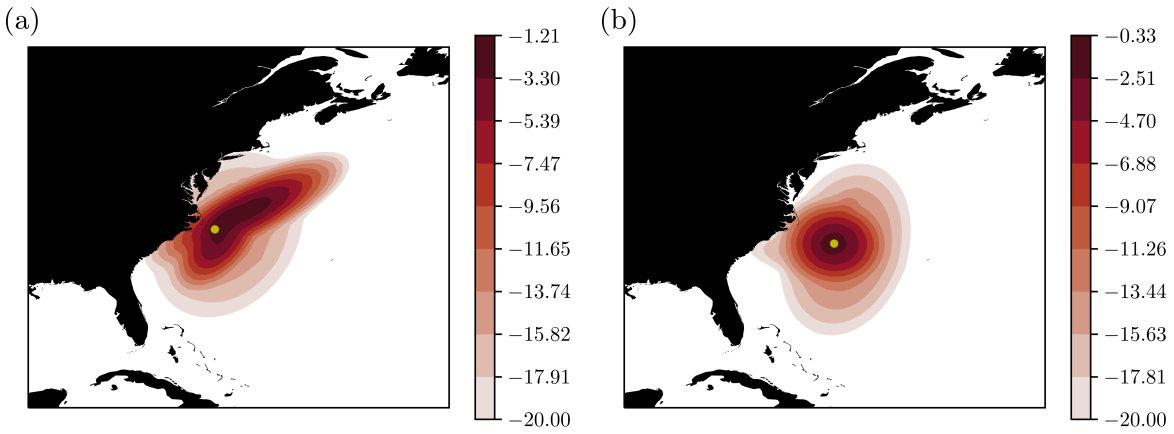


图 8：由混合密度网络模型推导的转移密度对数图，$\log\, p(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0)$，对应初始位置 $\boldsymbol{X}_0$，(a) 位于墨西哥湾流中（34.85° N， 74.50° W），以及 (b) 位于墨西哥湾流附近（33.67° N， 72.55° W），时间为 $\tau = 4$ 天。黄色圆点表示 $\boldsymbol{X}_0$。

在图 8 中，我们展示了两个不同初始位置 $\boldsymbol{X}_0$ 的对数转移密度 $\log\, p(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0)$。第一种情况下， $\boldsymbol{X}_0$ 位于墨西哥湾流核心区域内的 34.85° N、 74.50° W 处，转移密度呈现出明显的非高斯特征，其等值线大致向南方和东北方延伸，显示出墨西哥湾流对漂流浮标的影响。第二种情况下， $\boldsymbol{X}_0$ 位于墨西哥湾流外侧附近的 33.67° N、 72.55° W 处，转移密度更接近高斯分布。

为了量化转移密度偏离高斯分布的程度，以及这种偏离如何随海洋不同区域而变化，我们计算了单分量MDN模型（该模型为高斯分布）相对于完整32分量模型的库尔贝克–莱布勒（KL）散度[3]，并将其表示为初始位置的函数。结果如图 9 所示。请注意，由于两个高斯混合分布之间的KL散度尚无闭式表达式（Cui & Datcu 2015），我们在 $1° \times 1°$ 网格的每个顶点处基于5000个样本给出简单的蒙特卡洛估计。KL散度为零时，两个模型完全一致，表明位移服从高斯分布。KL散度越大，模型间的差异就越大，完整模型也就越偏离高斯分布。为帮助理解KL散度的大小，请注意，如果 $Z_0 \sim \mathrm{N}(m_0, 1)$ 且 $Z_1 \sim \mathrm{N}(m_1, 1)$，将它们的概率密度函数分别记作 p_0 和 p_1，则 $D_{\mathrm{KL}}(p_1 \parallel p_0) = (m_1 - m_0)^2$。位移的非高斯性可能主要源于海洋流速的不均匀性——漂流浮标在移动过程中会经过具有不同流场统计特性的区域，这些因素的综合效应会反映在观测位移中。另一种解释是，底层速度场本身具有非高斯性——Bracco等人（2000）和LaCasce（2005）基于海底流速计和海底漂流浮标的观测，提出了北大西洋海洋流速具有非高斯性的证据。

如图 8 所示，该模型为与陆地相交的漂流浮标位移赋予了非零概率。由于所假设参数形式（即高斯混合）的支撑集延伸至无穷远，这种情况无法避免；此外，这也未必完全是虚假的，因为确有一些漂流浮标会搁浅。2012年，Lumpkin 等人（2012年）重新评估了漂流浮标数据，以研究漂流浮标失效的原因。他们得出结论，约 27% 的漂流浮标

[3] p相对于 q的KL散度，也称为相对熵，定义为 $D_{\mathrm{KL}}(q \parallel p) = \int q(x) \log\, {}^{q(x)}\!/_{p(x)}\, \mathrm{d}x$，用于衡量概率密度 p 与参考概率密度 q 之间的差异——通常可解释为使用 p 近似 q时损失的信息量。

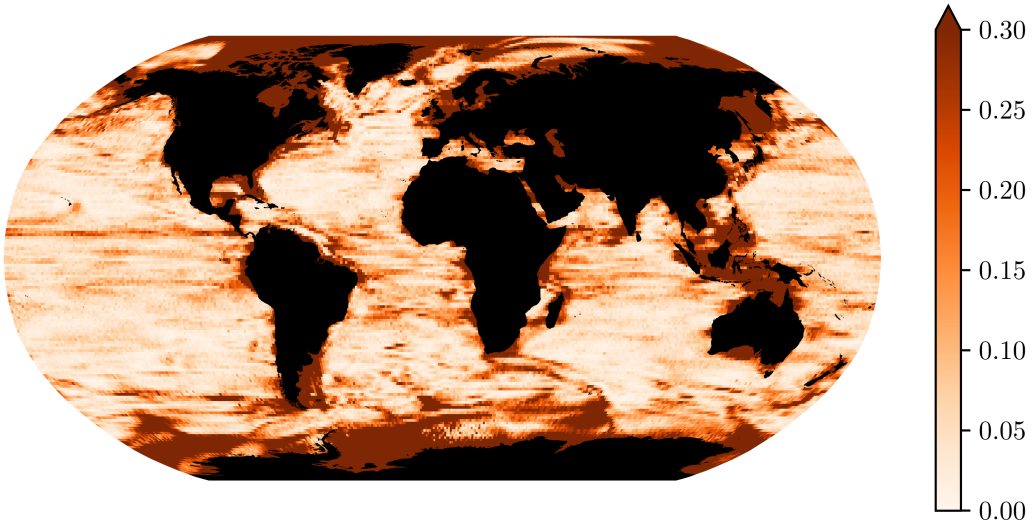


图9：单分量MDN模型与完整的32分量混合密度网络模型之间的 Kullback–Leibler 散度随初始位置的变化。数值越大，表示位移偏离高斯性的程度越强。

失效是由于搁浅，另有 10% 被人类捡走，其余则因内部故障而失效。在沿海区域以外，这一问题不太可能对第 4.4.1 节所讨论的位移统计量估计产生显著影响。第 4.4.2 节将进一步讨论其对漂流浮标模拟的影响。

4.4.1 位移统计量

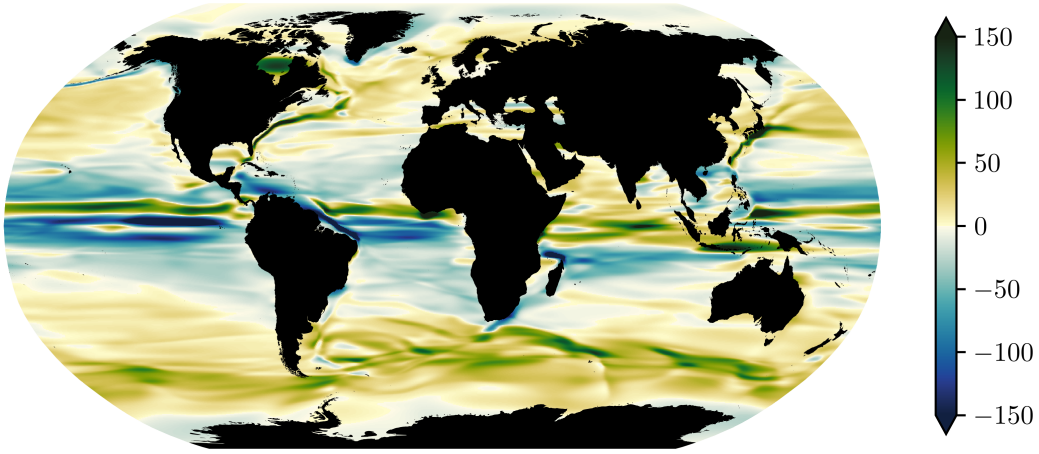
本节展示根据模型得出的单粒子统计量分布图。首先，我们以模型的4 天时间增量为例，展示位移的平均值。此外，我们还给出横向扩散率的全球估计。

图10展示了漂流浮标位移的平均值随初始位置的变化。尽管模型的输出采用经纬度坐标（λ，ϕ），我们仍基于局部切平面近似，进行简单换算，将其转换为公里

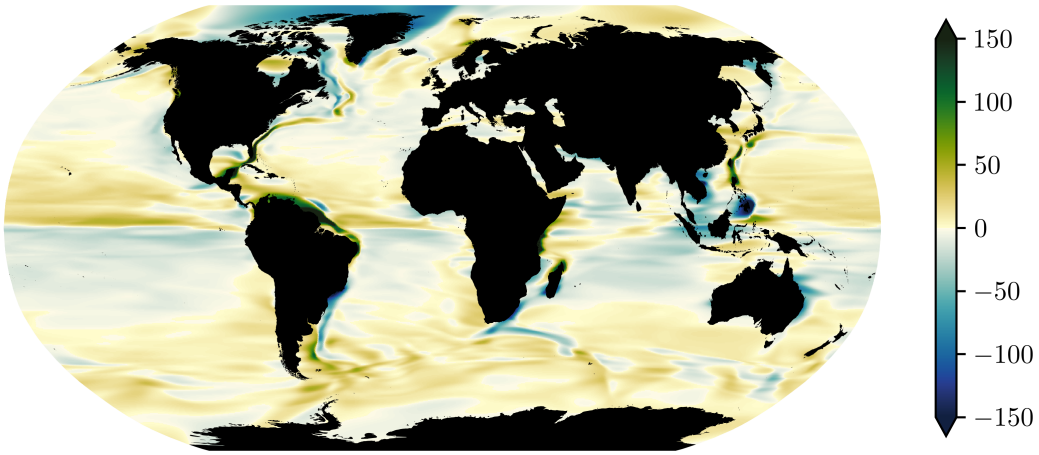
$$\Delta X = R\,\Delta\phi \tag{23a}$$
$$\Delta Y = R\,\Delta\lambda\,\cos\phi_0, \tag{23b}$$

其中 R 是地球赤道处的半径。海表动力学中的若干特征清晰可见，例如西边界流和赤道（逆）流。

为与以往研究进行比较，我们考虑利用模型估计横向扩散率，不过需要强调的是，通过对完整转移密度进行建模，我们对拉格朗日统计量的描述比 Davis（1987, 1991）提出的、用于描述扩散的经典平流-扩散模型更准确。多篇论文采用不同方法估计海洋扩散率（Oh et al. 2000, Zhurbas & Oh 2003, 2004, Klocker et al. 2012, Abernathey & Marshall 2013, Klocker & Abernathey 2014, Ying et al. 2019）。只有在存在合适的抽样时间时，才能直接根据漂流浮标位移估计扩散率；该时间必须长于漂流浮标速度的去相关时间，也就是说，要长于漂流浮标运动变为扩散运动所需的时间，而这样的



(a) 4天纬向位移的平均值（km）。



(b) 4天经向位移的平均值（km）。

图10：MDN模型在 $\tau = 4$天条件下的位移平均值随初始位置的变化。

即该时间尺度内的漂流浮标位移，与扩散率的空间变化相比，其尺度较小。在这种情况下，横向扩散率张量$\mathbf{K}(\boldsymbol{x})$可通过简单估算得到：

$$\mathbf{K}(\boldsymbol{x}) = \frac{1}{2\,\tau}\,\mathrm{Cov}(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0 = \boldsymbol{x}), \tag{24}$$

其中τ是经过适当选择的时间尺度，条件协方差则通过图1所示方法之一进行估算。遗憾的是，海洋中可能不存在这样的时间尺度；即便存在，它也可能随空间变化，因此难以确定。这一挑战已在以往的研究中得到证实（LaCasce 等（2014），Zhurbas 等（2014））。Oh 等（2000）提出了一种方法，以规避非均匀性带来的问题。他们建议分离出位移协方差中的横流分量，该分量由次主成分（即位移协方差中较小的特征值）确定，并据此得到扩散率的标量估计，因为横流分量受平均流切变的影响较小。图 11(a) 中给出了类似的估计，该估计由混合密度网络模型得出，时间尺度为 $\tau = 4$ 天，

$$K(\boldsymbol{x}) = \frac{1}{2\,\tau}\,\lambda_2(\boldsymbol{x}), \tag{25}$$

其中，$\lambda_2(\boldsymbol{x})$是.....的最小特征值

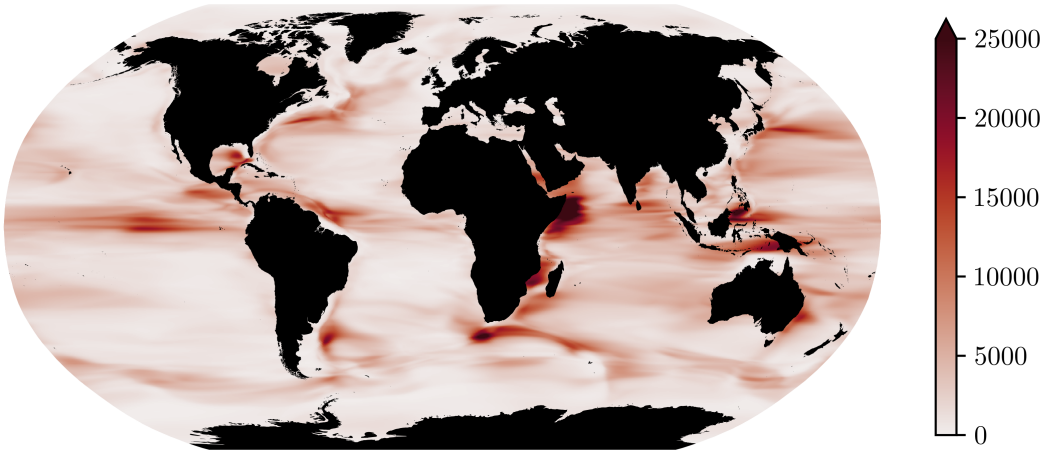
$$\mathrm{Cov}(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0 = \boldsymbol{x}) = \sum_i \alpha_i \mathbf{C}_i + \sum_i \left[\alpha_i \left(\boldsymbol{\mu}_i - \sum_i \alpha_i \boldsymbol{\mu}_i \right) \left(\boldsymbol{\mu}_i - \sum_i \alpha_i \boldsymbol{\mu}_i \right)^T \right]. \tag{26}$$

该结果与 Zhurbas 和 Oh（2004年）针对大西洋和太平洋给出的估计非常吻合。图 11(b)显示了分别采用$\tau = 14$ 天和 $\tau = 4$ 天时，按公式（25）得到的估计值之差。在许多区域，增大时间滞后 τ会使扩散率估计略有提高；在一些特别高能的区域，这种差异更加明显。不过，与基于位移协方差矩阵最大特征值推导出的类似沿流向扩散率估计相比，这种影响确实要弱得多。

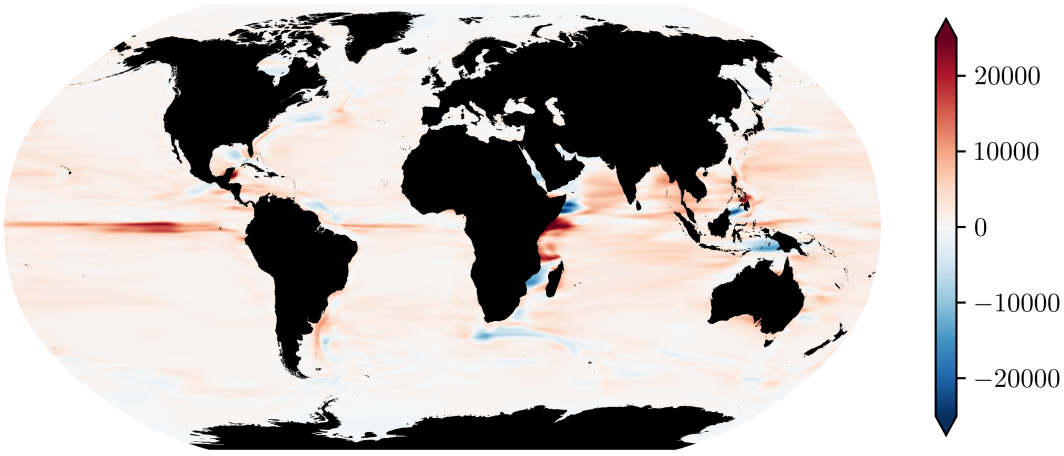
撇开从位移估计扩散率所面临的挑战不谈（这是所有方法共同面临的问题），我们在此着重说明本方法的一些优势。使用通过最大似然训练并借助早停实现有效正则化的混合密度网络（MDN）模型，消除了调整分箱分辨率的难题。我们的平均位移和扩散率估计的有效分辨率由数据分辨率自动决定，并能够在空间上进行最优调整。这样，我们便可一次性生成全局估计：在数据采样充分的区域，估计能够很好地解析流动特征；在数据相对稀疏的区域，估计也具有较好的适应性。但极高纬度区域除外，因为那里完全没有可用于约束模型的数据。

4.4.2 漂流浮标模拟

本节将演示如何以 MDN 模型为基础构建离散时间马尔可夫过程模型，并用它模拟漂流浮标轨迹。在离散时间设定下，假设具有马尔可夫性，意味着假设 $p(\boldsymbol{X}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{X}_{n-1} \cdots) = p(\boldsymbol{X}_{n+1} \mid \boldsymbol{X}_n)$。在这种情况下，抽样轨迹相当于按照转移密度依次反复抽样位移，因为给定当前位置后，位移在统计上独立于先前的位置。



(a) 根据训练天数为 $\tau = 4$ 天的混合密度网络模型推导出的横向扩散率标量估计值， $K^{(4)}$ ($m^2 s^{-1}$)。



(b) 分别根据训练天数为 τ =14 天和 $\tau = 4$ 天的混合密度网络模型推导出的横向扩散率标量估计值之差， K（14）− K（4）。

图11：根据转移密度的混合密度网络模型推导出的横向扩散率全球估计。

以这种方式模拟漂流浮标时，一个复杂之处在于，出于上文讨论的原因，漂流浮标可能会撞上陆地。在本研究中，我们不尝试对漂流浮标搁浅进行建模，因为目前尚不清楚全球漂流浮标计划数据集中是否包含足够可靠的信息——尤其是，判断漂流浮标是否搁浅仍是一项挑战（Lumpkin 等（2012））。为避免模拟中的漂流浮标搁浅，我们采用一种简单的拒绝采样方案：拒绝从转移密度中抽取的、会使漂流浮标到达陆地的位移，并重新抽取位移，直至抽到能使漂流浮标留在海洋中的位移。这相当于按照条件密度 $p(\Delta \boldsymbol{X} \mid \boldsymbol{X}_0,\ \boldsymbol{X}_0 + \Delta \boldsymbol{X} \notin \text{陆地})$进行抽样，也等同于使用转移矩阵模型时的标准做法，即将所考虑的定义域限制在海洋范围内，并相应地对概率估计进行归一化。为了判断拟议的新位置是否位于陆地上，我们检查其是否与分辨率为 110 米的陆地掩膜相交。

我们模拟了一组漂流浮标在 $2° \times 2°$ 网格顶点上的初始位置出发，历经 10 年的演变过程。请注意，每个漂流浮标的演变都是独立模拟的。这意味着，无法体现能够描述同时释放到海洋中的漂流浮标共同演变的多粒子统计量；尤其是，当前模型不适用于模拟短时间尺度内释放一团示踪粒子的情形。不过，可以预期，该模型能够描述漂流浮标或浮力示踪物在较大空间和时间尺度上的行为。Maximenko 等（2012）和 van Sebille 等（2012）也进行了类似实验，他们使用基于全球漂流浮标计划数据训练的转移矩阵模型，研究了模拟漂流器因近表层辐合而产生的聚集，以及所谓垃圾带的形成，包括北太平洋垃圾带（Moore 等（2001））和其他副热带海洋环流中的垃圾带。van Sebille 等（2012）的模拟还显示，巴伦支海中形成了另一个聚集区，但这一聚集区直到数十年后才出现。

我们的模型模拟结果总体上与这些先前研究一致。图12展示了混合密度网络模型下，模拟漂流器在模拟开始时以及经过一年、三年和十年演化后的分布。一年后，赤道地区的漂流浮标已变得相对稀疏。三年后，副热带环流中的簇开始显现；十年后，这些簇已十分清晰。此外，还出现了一些较小的簇，尤其是在北海和巴布亚以南的海域，以及一些高纬度区域，包括格陵兰西海岸沿岸和南极洲附近东经$100 - 130°$ 度一带。验证这些簇——也就是评估海洋垃圾是否可能在这些区域聚集——十分困难，因为原位观测仍然稀少（Ryan 等（2009））。这些区域的动力过程可能只是因 GDP 漂流浮标观测稀少而未被充分采样，导致混合密度网络模型对其刻画不足，进而产生虚假的辐合区。例如，我们在图10中注意到，平均位移似乎未能反映北海南部已知洋流的详细情况，而 GDP 数据中的漂流浮标并未到访该区域（见图2）。一般而言，与任何数据驱动模型一样，在解读数据缺失区域的模型输出时都应谨慎。

5 结论

本研究展示了条件密度估计，尤其是随机神经网络在一个流体动力学问题中的应用，即如何从单粒子

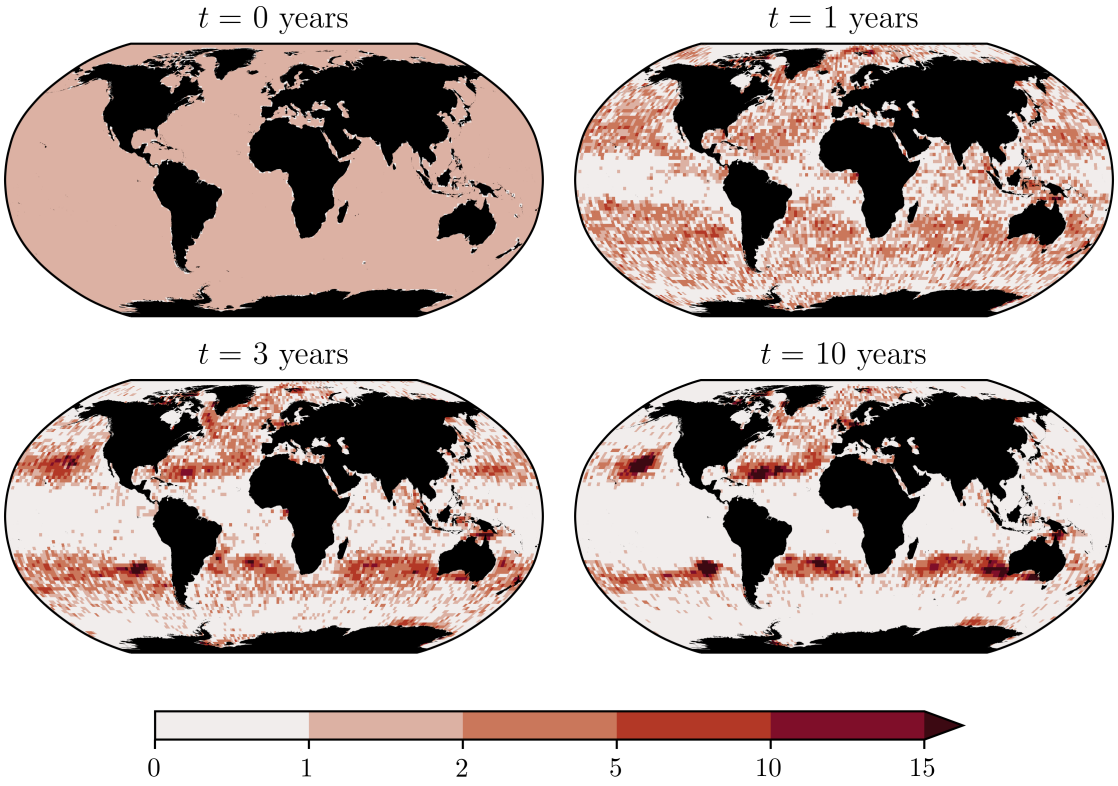


图12：在MDN模型下，模拟漂流器初始时以及演化一年、三年和十年后的直方图。

轨迹数据中诊断统计量。我们展示了这类概率模型如何既可用作模拟器，也可作为估计条件统计量的间接手段。采用概率建模框架，使我们能够借鉴统计推断、概率预测、模型比较和验证领域的大量文献，从而避免对损失函数和性能指标作出临时性的选择。我们使用概率评分规则将模型与替代模型进行比较，其中包括文献中广泛使用的马尔可夫链模型。结果表明，我们的模型无论在全局范围内还是在三个特定区域中，表现均优于这些模型。

通过对表层漂流器的单粒子转移密度进行建模，我们能够同时估算一系列条件统计量，从而捕捉海洋某些区域出现的强非高斯统计量。我们提供了平均位移和横向扩散率的全球分布图，但需要强调的是，这些示例仅对转移密度所包含的信息进行了有限概括；包括位移高阶矩在内的更多统计量，都可以轻易地通过我们的模型计算得出。将我们的模型视为离散时间马尔可夫过程的基础，我们还用它来模拟一组从均匀网格全球播撒的漂流浮标的演变，并展示了副热带环流中漂流浮标簇的形成，这与先前关于垃圾带形成的研究结果一致。

本文提出的方法同样适用于流体动力学和海洋学中的其他问题。例如，可根据欧拉或拉格朗日速度数据估计结构函数；也可以通过对多粒子转移密度建模，估计相对扩散等多粒子统计量；还可以学习气候、气象和海洋模型中的随机参数化。Guillau-

min 和 Zanna（2021）采用单分量MDN模型，在海洋模型中对亚网格动量强迫的参数化取得了进展；不过，这种方法适用范围更广，例如可用于亚网格输运的参数化。

在本研究中，我们基本忽略了量化模型参数不确定性以及在建模中纳入先验知识的需求。贝叶斯方法可以满足这些需求：我们不再通过最大似然估计参数，而是应用贝叶斯推断来获得参数的后验分布，并将先验知识纳入其中。事实上，本文所呈现的所有结果都将受益于不确定性量化。对于条件统计量，贝叶斯方法例如可以帮助识别哪些区域的数据不足以支持对横向扩散率作出可靠估计；一般而言，纳入先验知识也有助于对转移密度模型进行正则化，从而避免在漂流浮标模拟中出现虚假的辐合区。将贝叶斯推断应用于混合密度网络仍具挑战性，但我们认为这是未来一个重要的研究方向。

致谢。我衷心感谢 Jacques Vanneste、James Maddison 和 Aretha Teckentrup 提供的建议、意见以及对本研究的全力支持。我也感谢 Dhruv Balwada 参与有益的讨论。此外，还要感谢审稿人提出的建议，这些建议改进了本文。

经费支持。作者获得了英国工程与自然科学研究委员会（EPSRC）EP/S023291/1 项目资助的 MAC-MIGS博士培训中心支持。本研究使用了英国Cirrus国家二级高性能计算服务，该服务由爱丁堡并行计算中心（EPCC）提供（www.cirrus.ac.uk），经费由爱丁堡大学和英国工程与自然科学研究委员会（EP/S023291/1）资助。

利益冲突声明。作者声明不存在利益冲突。

数据可用性声明。复现本文结果所需的代码可在 doi.org/10.5281/zenodo.7737161 获取，其中还包含训练好的混合密度网络（MDN）模型以及演示其用法的 Jupyter笔记本。本文使用的处理后 GDP 数据和漂流浮标模拟数据可在 doi.org/10.7488/ds/3821 获取。

参考文献

Abernathey, R. P. 和 Marshall, J.（2013年），《由卫星测高推导的全球海表涡扩散率》，《地球物理研究杂志：海洋》**118**(2)，901–916。

Batchelor, G. K. (1953),《均匀湍流理论》，剑桥大学出版社。

Bernardo, J. M. (1979),‘作为期望效用的期望信息’，《统计年鉴》**7**(3), 686 – 690。

Bishop, C. M. (1994), 混合密度网络，技术报告 NCRG/94/004，阿斯顿大学。

Bishop, C. M. (2006),《模式识别与机器学习》，施普林格。

Bortolussi, L. & Palmieri, L. (2018),《化学反应网络的深度抽象》，载于M. ˇCeˇska & D. ˇSafr´anek, 编，《系统生物学中的计算方法》，施普林格国际出版，页码 21–38。

Bracco, A., LaCasce, J. H. & Provenzale, A. (2000), ‘海洋漂流浮标的速度概率密度函数’，《物理海洋学杂志》**30**(3), 461 – 474.

Br¨ocker, J. & Smith, L. A. (2007), ‘概率预测评分：恰当性的意义’,《天气与预报》**22**(2), 382–388. Cui, S. & Datcu, M. (2015), 高斯混合模型之间用于卫星图像检索的 Kullback–Leibler 散度近似方法比较，发表于 ‘2015 IEEE 国际地球科学与遥感研讨会（IGARSS）’，第 3719–3722 页。Davis, C. N., Hollingsworth, T. D., Caudron, Q. & Irvine, M. A. (2020), ‘混合密度网络在复杂流行病学个体模型仿真中的应用’,《公共科学图书馆·计算生物学》**16**(3), e1006869. Davis, R. E. (1987), ‘被动示踪剂的涡旋输运建模’,《海洋研究杂志》**45**(3), 635–666. Davis, R. E. (1991), ‘利用漂流浮标观测大尺度环流’,《深海研究*A*辑：海洋学研究论文》**38**, S531–S571. Du, H. (2021), ‘超越严格适当评分规则：局部性的意义’,《天气与预报》**36**(2), 457 – 468. Froyland, G. (2001), 通过马尔可夫模型提取动力学行为，发表于 ‘非线性动力学与统计学’，施普林格，第 281–321 页。Froyland, G., Stuart, R. M. & van Sebille, E. (2014), ‘全球海洋表面的连通性如何？’,《混沌：非线性科学跨学科期刊》**24**(3), 033126. Gneiting, T. & Raftery, A. E. (2007), ‘严格适当评分规则、预测与估计’,《美国统计学会杂志》**102**(477), 359–378. Guillaumin, A. P. & Zanna, L. (2021), ‘海洋动量强迫的随机深度学习参数化’,《地球系统建模进展》**13**(9), e2021MS002534. Hansen, D. V. & Herman, A. (1989), ‘热带太平洋海面漂流浮标的时间采样要求’,《大气与海洋技术杂志》**6**(4), 599–607. Hansen, D. V. & Poulain, P.-M. (1996), ‘WOCE-TOGA漂流浮标数据的质量控制与插值’,《大气与海洋技术杂志》**13**(4), 900–909. Kingma, D. P. & Ba, J. (2015), ‘Adam：一种随机优化方法’, *arXiv:1412.6980*. Klocker, A. & Abernathey, R. (2014), ‘中尺度涡旋特性与扩散率的全球分布模式’,《物理海洋学杂志》**44**(3), 1030–1046. Klocker, A., Ferrari, R., LaCasce, J. H. & Merrifield, S. T. (2012), ‘协调基于漂流浮标与基于示踪剂的横向扩散率估计’,《海洋研究杂志》**70**(4), 569–602. LaCasce, J. (2005), ‘北大西洋的欧拉速度分布与拉格朗日速度分布’,《物理海洋学杂志》**35**(12), 2327–2336. LaCasce, J. (2008), ‘拉格朗日观测的统计特征’,《海洋学进展》**77**(1), 1–29.

LaCasce, J. H., Ferrari, R., Marshall, J., Tulloch, R., Balwada, D. & Speer, K. (2014), ‘Dimes实验中基于浮标的等密度面扩散率’, 《物理海洋学杂志》**44**(2), 764 – 780. LeCun, Y. A., Bottou, L., Orr, G. B. & M¨uller, K.-R. (2012), 高效反向传播, 施普林格, 柏林, 海德堡, pp. 9–48. Liu, Z., Hartwig, T. & Ueda, M. (2020), ‘神经网络无法学习周期函数及其解决方法’,*arXiv:2006.08195* . Lumpkin, R. & Centurioni, L. (2019), ‘数据集：全球漂流浮标计划的海洋表层漂流浮标经质量控制的6小时插值数据’. 访问日期：2022年12月4日。Lumpkin, R., Maximenko, N. & Pazos, M. (2012), ‘评估漂流浮标在哪里以及为何失效’,《大气与海洋技术杂志》**29**(2), 300 – 308. Maulik, R., Fukami, K., Ramachandra, N., Fukagata, K. & Taira, K. (2020), ‘用于流体流动代理建模和数据恢复的概率神经网络’, 《物理评论：流体》**5**(10), 104401. Maximenko, N., Hafner, J. & Niiler, P. (2012), ‘由拉格朗日漂流浮标轨迹推导出的海洋垃圾迁移路径’, 《海洋污染通报》**65**(1), 51–62. Miron, P., Beron-Vera, F. J., Helfmann, L. & Koltai, P. (2021), ‘海洋垃圾的转变路径与垃圾带的稳定性’, 《混沌：非线性科学跨学科期刊》**31**(3), 033101. Miron, P., Beron-Vera, F. J., Olascoaga, M. J., S heinbaum, J., P´erez-Brunius, P. & Froyland, G. (2017), ‘墨西哥湾的拉格朗日动力地理学’, 《自然-科学报告》**7**(1), 1–12. Monin, A. S. & Yaglom, A. M. (1971), 统计流体力学：湍流力学, 第1–2卷, 麻省理工学院出版社。Moore, C., Moore, S., Leecaster, M. & Weisberg, S. (2001), ‘北太平洋中央环流中的塑料与浮游生物比较’, 《海洋污染通报》**42**(12), 1297–1300. Oh, I. S., Zhurbas, V. & Park, W. (2000), ‘根据卫星跟踪漂流浮标数据估算东海（日本海）和西北太平洋的水平扩散率’, 《地球物理研究杂志：海洋》**105**(C3), 6483–6492. Pavliotis, G. A. (2014), 随机过程及其应用：扩散过程、福克–普朗克方程和朗之万方程, 第60卷，收录于《应用数学丛书》，施普林格，纽约。Prechelt, L. (2012), 提前停止——何时为宜？, 施普林格, 柏林, pp. 53–67. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1985), 通过误差传播学习内部表征，技术报告，加州大学圣迭戈分校认知科学研究所。Rupolo, V. (2007), ‘用于确定轨迹分类和湍流状态的基于拉格朗日的方法’,《物理海洋学杂志》**37**(6), 1584 – 1609.

Ryan, P. G., Moore, C. J., Van Franeker, J. A. & Moloney, C. L. (2009), ‘监测海洋环境中塑料碎片的丰度’, 英国皇家学会哲学汇刊*B*辑：生物科学**364**(1526), 1999–2012. Shin, J., Ge, Y., Lampmann, A. & Pfitzner, M. (2021), ‘湍流预混燃烧大涡模拟中的数据驱动亚网格尺度模型’, 《燃烧与火焰》**231**, 111486. TensorFlow开发者 (2021), ‘TensorFlow’. v2.7.0, doi. TensorFlow Probability开发者 (2021), ‘TensorFlow Probability’. v0.15.0, GitHub发行版。Ulam, S. M. (1960),《数学问题集》, Interscience Publishers. van Sebille, E., England, M. H. & Froyland, G. (2012), ‘基于观测表层漂流浮标的海洋垃圾带起源、动力学与演变’, 《环境研究快报》**7**(4), 044040. van Sebille, E. et al. (2018), ‘拉格朗日海洋分析：基础与实践’, 《海洋建模》**121**, 49–75. Ying, Y. K., Maddison, J. R. & Vanneste, J. (2019), ‘基于拉格朗日轨迹数据的海洋扩散率贝叶斯推断’,《海洋建模》**140**, 101401. Zhurbas, V., Lyzhkov, D. & Kuzmina, N. (2014), ‘基于漂流浮标的世界海洋横向涡扩散率估算，重点关注印度洋及参数化问题’, 《深海研究第一辑：海洋学研究论文》**83**, 1–11. Zhurbas, V. & Oh, I. S. (2003), ‘根据漂流浮标数据推导的太平洋横向扩散率与拉格朗日尺度’, 《地球物理研究杂志：海洋》**108**(C5). Zhurbas, V. & Oh, I. S. (2004), ‘与表层环流模式相关的太平洋和大西洋横向扩散率漂流浮标推导图’, 《地球物理研究杂志：海洋》**109**(C5).