arXiv:2305.03214v1 [stat.AP] 5 May 2023

原创研究文章

使用状态空间模型分析生态瞬时评估数据：注意事项与建议

Lindley R. Slipetz 美国弗吉尼亚大学，弗吉尼亚州夏洛茨维尔，美国 Jeremy W. Eberle 美国弗吉尼亚大学，弗吉尼亚州夏洛茨维尔，美国 Cheri A. Levinson 美国路易斯维尔大学，美国肯塔基州路易斯维尔 Ami Falk 美国弗吉尼亚大学，弗吉尼亚州夏洛茨维尔，美国 Claire E. Cusack 美国路易斯维尔大学，美国肯塔基州路易斯维尔 Teague R. Henry 美国弗吉尼亚大学，弗吉尼亚州夏洛茨维尔，美国

文章历史记录
编译于2023年5月8日

摘要
生态瞬时评估（EMA）数据广泛应用于时间趋势和关系的研究。在生态瞬时评估研究中，有许多设计方面的考量会影响数据分析。有一种通用的建模框架尤其适用于此类分析：状态空间建模。本文介绍状态空间建模框架，并就生态瞬时评估数据建模所涉及的考量提出建议。这些建议可应对生态瞬时评估数据分析中出现的问题，例如个体化与普遍化建模、缺失数据，以及平稳与非平稳数据。此外，我们还推荐了可用于实践这些建议的 R 软件包。总体而言，设计良好的生态瞬时评估研究为研究者提供了机会，使其能够在评估心理现象时把握瞬息万变的细微之处。

关键词
状态空间模型；生态瞬时评估

生态瞬时评估（EMA）是一种时间序列收集技术，通常在数周内、参与者所处的自然环境中每天多次收集测量数据。纵向数据有多种子类型，包括面板数据和密集纵向数据。传统上，面板数据的样本量较大，而时间点数量较少；生态瞬时评估最恰当地应被描述为后者。由于生态瞬时评估数据具有复杂性，即测量值之间存在依赖、缺失，以及其他设计与建模决策点，在这一媒介中进行正确建模是一项挑战。在本文中，我们考察这些研究者决策点，并针对研究生态瞬时评估数据的应用研究者所面临的障碍，提出设计与建模方案。通过采用状态空间模型这一通用建模框架，我们展示了如何对这些选择逐一进行定量处理，例如，研究者应根据数据收集时间的变异性选择何种模型。总体而言，本文旨在指导研究者使用生态瞬时评估和状态空间建模。

与生态瞬时评估研究不同，实验室内实验或评估只能提供所研究现象的一个截面，无法考虑评估前数日或数周内的变化（Wright et al., 2021）。此类评估还依赖参与者的记忆和/或其对既往经历进行概括的能力（Mote & Fulford, 2020）。例如，在临床访谈中，回答“在过去两周内，当你感到抑郁时是否感到疲劳？”这一问题，需要参与者回忆过去两周内抑郁和疲劳出现的情况，并判断在这段时间内二者是否存在总体关系。此外，实验室内评估还存在情境影响这一问题，其中信任、文化偏差等因素会影响参与者披露信息时的舒适程度（Wright et al., 2021）。生态瞬时评估数据为实验室内评估方法提供了一种替代方案，可以克服这些障碍。

从本质上说，生态瞬时评估是一种数据收集方案，用于在参与者的生活体验中评估其思想、感受、行为，甚至生理状况。参与者在进行日常生活时，其回应会以实时或近实时的方式收集，通常通过参与者智能手机上的应用程序完成。这种设计最大限度地减少了对回忆的需求，并能够收集精确的时间信息。从理论上讲，与实验室内评估相比，生态瞬时评估的偏倚较小；例如，它减少了对多次经历进行回忆或概括的需要（Mote & Fulford, 2020）。例如，在生态瞬时评估研究中，参与者不必回答“在过去两周内，当你感到抑郁时是否感到疲劳？”，而是会在数周内每天多次被询问其疲劳和抑郁情况。

生态瞬时评估的收集方式有两种：设备提示式（例如，数据收集设备会在一天中的特定时间提示用户回答问题），或参与者发起式（例如，发生某个事件后，参与者将其记录下来）（Srivastava, Michael, Manasse, & Juarascio, 2021）。例如，可以要求参与者记录特定行为的发生情况（例如，吸烟或饮酒时），并提示他们提供更多信息（例如当前的情感状态）；也可以在一天中多次提示（即“呼叫”）参与者作答，并根据一天中的不同时间安排不同的问卷（例如，白天进行简短的情感状态问卷，在晚间提示时完成较长的一组题目）。另一种参与者发起式安排是地点发起式：参与者到达指定地点时（例如，在一项酒精使用研究中，到达酒吧时），可能会被要求作答。除参与者发起式和按计划进行的数据收集外，如果还要开展移动心理生理学研究，则可能需要昼夜持续进行生理监测。

收集完成后，生态瞬时评估数据为分析时间趋势和变量间关系提供了丰富的基础。例如，在一项酒精使用研究中，研究者可能会关注酒精使用与一天中的时段或一周中的星期几之间的关系。研究者采用的方法不同，对生态瞬时评估数据进行分析所得的结果，所能聚焦推断的数据方面也会大不相同。生态瞬时评估数据可能的用途和分析方式，可通过以下示例一窥究竟：

Ditzen、Hoppmann 和 Klumb（2008年）研究了亲密关系与每日唾液皮质醇水平之间的关系，以评估伴侣关系中的健康保护因素。他们的EMA研究采用自我报告的方式，记录参与者用于亲密关系的时间、工作中的长期问题以及情感质量；同时，研究人员每三小时收集一次唾液样本，持续一周，以估算皮质醇水平。值得注意的是，该EMA方案将现代生态瞬时评估研究中常见的通过智能手机进行的行为/心理测量，与参与者提供的唾液样本结合起来。尤其是，唾液样本的采样率与行为测量相同，因此可以在相同的时间尺度上分析皮质醇与行为之间的关系。尽管实际收集唾液样本会增加参与者负担，这项研究表明，在获取更难收集的数据类型的同时，也可以采用类似EMA的方案。

MacKerron 和 Mourato（2013年）旨在评估幸福感与物理环境之间的联系。在这项EMA研究中，参与者会在自己选择的时间区间和频率内随机收到提示。他们需要在连续滑动刻度上报告自己的幸福程度、当时与谁在一起、所在地点以及正在做什么。填写问卷时，研究人员通过GPS获取了参与者的精确位置。随后，GPS数据与三个指标相关联：栖息地/土地类型（例如绿地）、天气状况和日照状态。这项研究因其规模以及对GPS位置数据的使用而值得关注。虽然生态瞬时评估数据收集通常与较小的样本量相关，但该研究表明，EMA也可以大规模开展：该研究从21,947名英国参与者那里获得了1,138,481个数据点。它还展示了被动数据收集的应用：通过收集GPS坐标，研究人员能够研究情感与位置之间的关联，而无需依赖参与者可能不可靠的事后位置自我报告。

生态瞬时评估方案还可以与实验方法结合，例如考察干预对行为瞬时动态的影响。在 Levinson 等人（2023年）最近开展的一项临床试验中，研究者通过生态瞬时评估数据收集来识别进食障碍中因人而异的核心症状，并据此制定为期数周的个性化治疗计划。在该研究中，生态瞬时评估数据是在干预前收集的，而干预效果则通过更为传统的单次测量问卷进行评估。总的来说，除了参与者负担之外，没有其他障碍会妨碍在干预前、干预期间和干预后收集生态瞬时评估数据，从而近乎实时地追踪行为动态的变化。生态瞬时评估数据收集本身也可以成为干预的一部分。这类生态瞬时评估方案被称为生态瞬时干预（EMI；McDevitt-Murphy、Luciano 和 Zakarian，2018年），并已显示出治疗多种心理障碍／行为问题的潜力，其中尤以抑郁症为甚（Schueller、Aguilera 和 Mohr，2017年）。例如，在 Levinson 等人（2023年）的研究中，如果发现某位参与者的核心认知症状是害怕被拒绝，那么干预措施就会是社会暴露，随后进一步收集生态瞬时评估数据，并针对其他核心症状实施干预。

总之，与实验室内实验或面板纵向研究相比，生态瞬时评估有两个关键优势。首先，生态效度得到了提升。

效度；也就是说，通过生态瞬时评估数据，参与者在自然环境中的行为和反应得以被记录。其次，生态瞬时评估数据提供了细粒度的时间信息，使研究人员能够分析行为或心理状态的日内动态。这是传统的面板纵向设计无法做到的。

然而，尽管生态瞬时评估数据前景广阔，但研究者如何分析这类数据，仍受到许多设计和分析方面因素的影响。本文首先回顾一种通用建模框架——状态空间建模，它尤其适用于生态瞬时评估数据的分析。状态空间建模是一种动态潜变量模型，其中未观测状态变量追踪所研究现象的发展轨迹。接下来，我们将介绍生态瞬时评估研究及其数据方面的若干注意事项、这些事项对分析和推断的影响，以及可用于处理这些问题的相关方法。

1. 用于生态瞬时评估数据的状态空间模型

状态空间建模框架用于时间序列数据分析，是一种通用框架，它将随时间变化的观测变量表示为一组潜在、未观测状态 变量的测量值（Kalman，1960）。状态空间模型将经典心理测量（即验证性因素分析，或 CFA）与未观测潜在状态的时间动态结合起来。例如，在抑郁症的 CFA 模型中，潜变量“抑郁症”由抑郁问卷中选取的一组条目进行测量。将这一模型扩展到时间维度后，一项 EMA 研究可以收集同一组抑郁问卷条目的个体内重复测量值，此时，抑郁症将成为一个潜在状态，由问卷回答在一段时间内进行测量。状态空间建模框架既能让研究者对抑郁症的测量加以考量，也能评估抑郁症随时间与自身的关联。

作为将状态空间建模应用于心理学数据的一个示例，McKee、Russell、Mennis、Mason 和 Neale（2020年）提出了一种情绪稳态反馈控制模型，旨在为一名高风险青少年（即13至14岁）样本在18个时间点上的11个情感条目时间序列建立情绪调节指标。在这个连续时间模型中，状态是情感，通过观察到的情感测量值来衡量；情绪调节指标则是自回归调节参数（定义为每个时间段内恢复到均衡状态的百分比），以及外生扰动的标准差（定义为个体身上未测量到的影响）。研究者对尼古丁、酒精和大麻使用总次数进行横断面个体间分析，并以调节参数、外生扰动的标准差，以及各测量指标总分的均值/标准差作为预测变量。他们发现情感状态与物质使用之间存在显著关系，这或许表明存在双向反馈循环。例如，较高的负性情感与更强的吸烟倾向相关，而较小的情绪标准差也与更强的吸烟倾向相关。这项研究表明，状态空间模型能够追踪情感调节随时间的变化，具体而言，也能够一般性地追踪心理现象的调节随时间的变化。

Perh或许最简单、最常用且研究最多的状态空间模型是离散

时间单个参与者时间序列的多元正态状态空间模型:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t + \boldsymbol{\varepsilon}_t \tag{1}$$
$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \boldsymbol{\nu}_t \tag{2}$$
$$\boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma}) \tag{3}$$
$$\boldsymbol{\nu}_t \sim N(0, \boldsymbol{\Theta}) \tag{4}$$

其中， x_t 是在时间 t 的未观测状态（例如抑郁症）的值向量，u_t 是观测到的外生扰动的向量（例如外部事件，如家人去世，或上文提到的社会暴露等干预）， y_t 是在时间 t 的观测测量值向量（例如贝克抑郁量表中的条目）， **A** 是一个方阵，其系数描述状态值如何随时间相互关联（例如时间 t 与 $t+1$ 时抑郁症之间的自回归关系），**G** 是一个矩阵，用于将外生扰动的影响（例如预定干预的影响）映射到各状态上；**H** 是一个系数矩阵，描述每个观测指标如何映射到潜在状态值上（即类似 CFA 中的因子载荷）。 $\boldsymbol{\varepsilon}_t$ 和 $\boldsymbol{\nu}_t$ 分别是状态创新（即与状态相关的随时间变化的变异性）和测量误差变量，其分布均为相应维度的多元正态分布，期望值为 0，协方差矩阵分别为 $\boldsymbol{\Sigma}$和 $\boldsymbol{\Theta}$ 。

细心的读者会注意到，方程 1-4 中描述的模型与经典向量自回归（即向量自回归）模型以及心理学和行为科学领域常用的潜变量模型都很相似。状态空间框架具有很强的通用性，许多标准模型都可以用状态空间表示法来描述。例如，将 x_t 视为观测变量，并完全忽略 y ，就可以把上述模型简化为一阶向量自回归模型（即模型只定义在一个滞后阶数上）。就其与潜变量模型的相似之处而言，状态变量 x 是潜变量，y 是显变量指标，**H** 矩阵可视为因子载荷矩阵（在结构方程模型文献中通常记作 $\boldsymbol{\Lambda}$ ），**A** 与因子间回归系数矩阵（通常记作 **B**）类似，而 $\boldsymbol{\varepsilon}_t$、 $\boldsymbol{\nu}_t$ 分别是潜变量和指标的误差分布。大多数用于表示变量时间动态的模型都可以写成状态空间模型。除测量这一需要假设存在潜在状态的情形外，本文讨论的问题和注意事项均不取决于潜在状态是否存在，并且适用于更广泛的多元时间序列模型，例如向量自回归模型。尽管如此，我们仍聚焦于状态空间框架，因为它具有通用性，且适用于范围广泛的生态瞬时评估数据。

还必须注意，公式 1-4 所描述的模型包含若干强假设，但这些假设可以放宽或更改，并非状态空间建模框架所固有的。例如，该模型以离散时间运行，即认为从 t 到 $t+1$ 的每个时间区间都相同。在离散模型中，测量值之间的滞后阶数的重要之处不在于每个滞后的长度，而在于每个滞后都相同。该模型还是时不变的，即各种模型参数在整个研究期间都不会变化。此外，该模型描述的是单个多变量时间序列的动态，而非多个参与者的时间序列数据所呈现的总体动态。最后，该模型假设状态和观测都符合多元正态分布，这也意味着所有时间动态之间的关系都.....

状态可以用线性方式描述。同样，这些假设都不是状态空间建模框架固有的。在本文余下部分中，我们将展示如何修改上述模型（即放宽或改变假设），以适应生态瞬时评估数据的各种特征。

本文其余部分的结构安排如下：我们将介绍若干注意事项，我们认为计划开展生态瞬时评估研究或分析生态瞬时评估数据的科学家应当熟悉这些事项，并在其设计/分析中予以考虑。对于每一项注意事项，我们将讨论其具体内容、其为何对生态瞬时评估数据的分析至关重要，以及在可能的情况下，如何修改公式 1–4 中描述的模型，以考虑该注意事项。虽然我们不会详细介绍这些模型的估计方法，但我们将提供适用的 R 软件包参考资料，以及有关各种状态空间建模方法的延伸阅读参考文献。请注意，我们明确且仅引用公开可用的 R 软件包，既是为了缩小综述范围，也是为了鼓励使用开源且免费提供的科学软件。

2. 注意事项与建议

2.1. 个体化与普遍化建模

生态瞬时评估会收集丰富的参与者内数据，因此每位参与者通常会产生大量观测值。研究人员在权衡分析策略时应考虑的第一个建模方面是：他们认为所研究的现象在不同参与者之间会有很大差异、在所有参与者之间都相同，还是介于这两个极端之间。这就是个体化与普遍化建模的考量（Silverstein，1988年）。

考虑一项 EMA 研究，其中正在对抑郁症（Fried & Nesse，2015）这种高度异质性障碍的动态进行建模。纯粹的个体化方法意味着每个人都有一套独特的症状动态；部分个体化方法意味着样本中存在某种共同的动态模式，但同时也存在个体间变异；而完全的普遍规律取向则意味着所有个体的抑郁症状都由完全相同的动态所驱动。显然，选择个体化方法还是普遍规律取向，与研究现象背后的理论密切相关。

2.1.1. 纯个体化与普遍化方法

当现象完全或部分具有个体化特征时，错误地采用普遍规律取向会带来风险；当普遍规律取向更为合适时，错误地采用个体化方法也会带来风险。在前一种情况下，普遍化模型很可能无法很好地代表任何一位参与者；在纯粹个体化过程这一最糟糕的情况下，甚至无法代表任何参与者的动态（Mole-naar，2004年）。另一方面，在后一种情况下，如果所有个体的现象动态都相同，却采用个体化方法，就会因未能汇总个体间的信息而导致检验效能损失和估计偏倚（Gelman，2006年）。鉴于在普遍规律情形下，动态很可能无法得到恰当表征，我们建议研究者倾向于采用个体化分析，除非有充分的理论理由相信该现象具有普遍规律性。

当纯粹的个体化方法适用与纯粹的普遍规律取向适用这两个极端情形，在状态-

空间建模框架中，纯粹采用个体化方法与纯粹采用普遍规律取向各自适用的两种极端情形，均易于表示。对于纯粹的个体化方法，可分别对每个人的时间序列建模，从而为每个人估计出一组独特的模型矩阵（$\mathbf{A}, \mathbf{G}, \mathbf{H}$ $\mathbf{\Sigma}, \mathbf{\Theta}$）。类似地，在纯粹的总体化情形中，可以将个体时间序列拼接起来（并用间隔表示参与者数据之间的分界），然后拟合单个模型，估计出一组模型矩阵；理论上，这组矩阵能够代表样本中每位参与者的动态。使用常见的状态空间建模软件包（如 dlm（Petris，2010年）或 dynr（Ou、Hunter 和 Chow，2019年））即可较为轻松地估计这两种情形。这两个软件包都能够使用卡尔曼滤波器拟合离散时间状态空间模型（参见“缺失数据”部分）。

2.1.2. 部分个体化方法

部分个体化方法则更为复杂。在这种方法中，所有个体共享一组动态，但允许个体在群体动态的基础上有所变化。这里的一个问题是，个体围绕群体动态的变化这一概念定义不明确。例如，个体可能具有相同的一组结构参数（即**A** 矩阵中的非零元素），但这些非零参数可以取不同的值。另一种情况是，整个样本共享一组关系（即抑郁症状态与焦虑状态之间的交叉滞后关系），同时每个个体还具有一组额外关系（即部分个体的焦虑与冲动性之间可能存在关系，但并非所有纳入分析的个体都如此）。

有几种模型/软件包可用于对多变量时间序列数据进行部分个体化建模。gimme（Lane 等，2022）软件包实现了群组迭代多模型估计算法（Gates 与 Molenaar，2012年），该算法通过迭代方式将结构向量自回归模型拟合到多个时间序列，以估计群体路径（即**A** 矩阵中整个样本共有的动态）和个体路径。逐步模型搜索程序会分别为群体路径和个体路径估计稀疏动态。研究者在估计包含潜变量的部分个体化模型（即传统状态空间模型），或关注亚组的自动检测与验证性检验时，应使用 gimme 软件包（Henry 等，2018；Lane、Gates、Pike、Beltz 与 Wright，2018年）；不过，该软件包也能估计仅包含观测变量的模型。

或者，multivar 软件包（Z. Fisher，2022年）提供了一种针对观测变量的多参与者正则化向量自回归模型的实现。该方法使研究人员能够估计群体效应和个体效应，同时结合正则化的能力来估计稀疏动态（即，A 矩阵中的某些元素被设为 0）。在拟合具有大量参数的观测变量多参与者模型时，研究人员应使用 multivar 软件包（因为正则化会将部分参数收缩至零，从而使模型的计算更加简单，并提高其可解释性）。GIMME 和 multivar 均可估计稀疏动态，不过，GIMME 通过模型构建方法来估计稀疏动态，由此得到群体层面效应以及个体特异性效应。

mlVAR（Epskamp、Deserno 和 Bringmann，2021）软件包实现了一种多层或混合效应向量自回归模型，利用随机效应对观测变量动态中的个体间变异进行建模。多层向量自回归不会产生稀疏动态，但会为所有参数（指定了随机效应的参数）产生群体层面效应和个体层面效应。变量较少且希望估计多参与者观测变量模型的研究人员应使用 mlVAR。需要指出的是，除 gimme 之外，没有其他 R 软件包可用于估计包含潜在状态的部分个体化模型，

这些模型所针对的多变量时间序列部分个体化建模是一个活跃的研究领域，尤其是在健康和行为科学中，研究人员经常收集多名参与者的时间序列数据，目前也在积极开发新的创新方法。关于 EMA 情境下个体化与普遍化建模的入门读物，我们推荐 A. J. Fisher、Medaglia 和 Jeronimus（2018）；Foster 和 Beltz（2018）；Molenaar 和 Beltz（2020）。

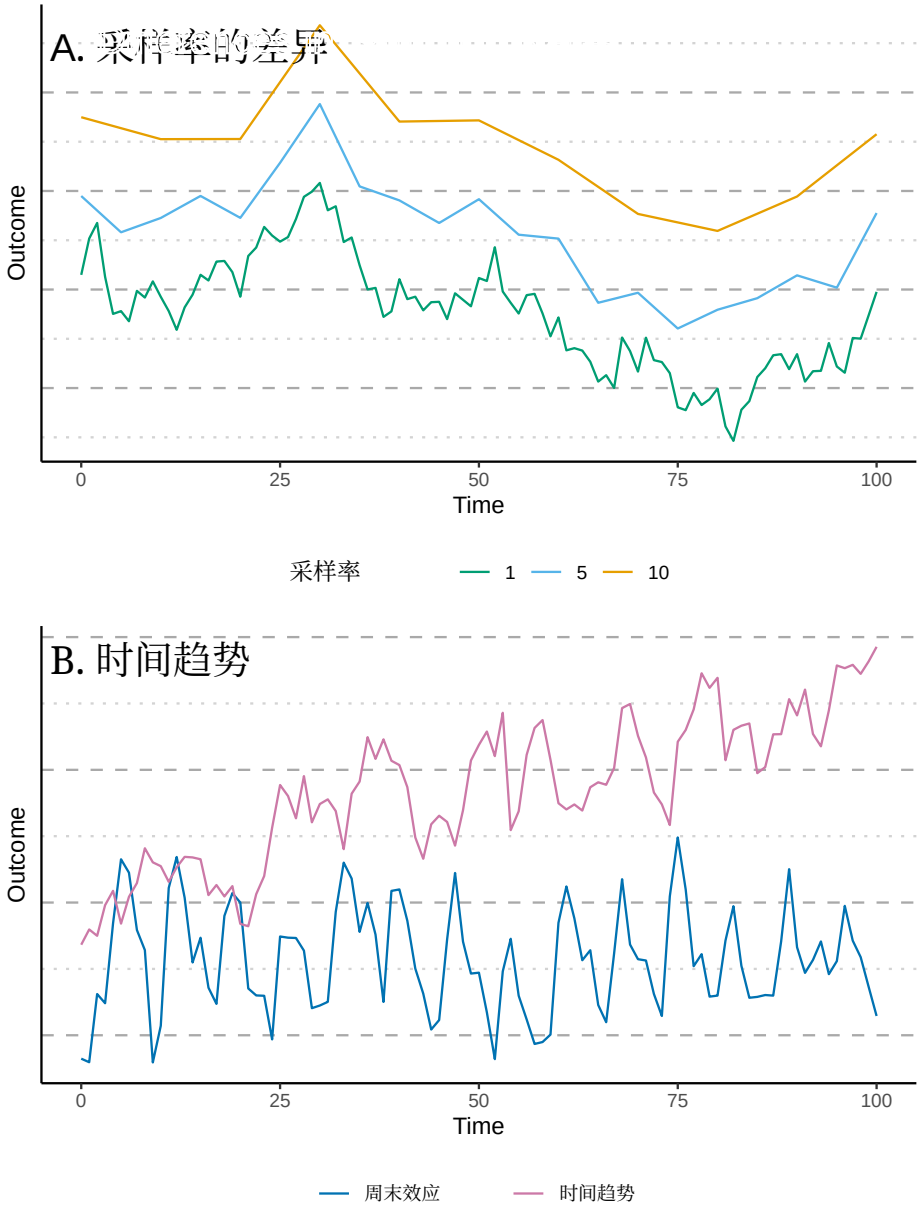

图1. 面板 A 展示了采样率与动态模式之间的关系。绿色线表示足以捕捉动态模式的采样率。蓝色线将采样率降低至原始数据中每 5 个时间点取 1 个时间点。请注意，蓝色线无法捕捉较小的波动，但仍遵循整体模式。橙色线表示进一步降低采样率（原始数据中每 10 个时间点取 1 个时间点）。进一步降低采样率会削弱捕捉动态模式的能力。面板 B 展示了两种时间趋势：随时间线性增加（洋红色线）和周末效应（蓝色线）。在这张图中，时间序列的参数相同，只有时间趋势不同。这表明时间趋势会对时间序列产生显著影响。

2.2. 数据收集时间表

有若干研究设计和分析方面的考量可能会影响提示时间表，即在研究期间的数天、数周或数月内何时收集数据。提示时间表大致分为两类（也可以组合使用）。第一类是预定提示，即在特定时间向参与者发送回应提示，或告知参与者在一天或一周中的特定时间记录数据（例如每天中午）。第二类是事件驱动提示，即在满足特定条件时提示参与者作出回应（例如，GPS测得参与者进入某个地点时）；更常见的是，要求参与者在特定行为或事件发生之前、期间或之后作出回应（例如，在一项研究酒精使用的研究中，要求参与者在一次饮酒后作出回应；或在一项饮食失调研究中，要求参与者在用餐后作出回应）。

2.2.1. 采样率

在研究设计和分析层面，有若干与提示时间表相关的考量。最重要的是，数据收集必须足够频繁，才能测量理论上相关过程的变化。图1的A面板展示了降低采样率如何掩盖动态中的相关模式。绿色曲线可视为以足以捕捉所有动态的速率采集的数据。蓝色曲线将采样率降低至原始数据每5个时间点采集1个时间点。请注意，采样率降低后，较小的波动会被掩盖，但总体模式仍得以保留。进一步降低采样率（黄色曲线）会使我们追踪动态波动的能力日益下降。在设计生态瞬时评估（EMA）研究时，研究人员应谨慎确定适合所研究现象的采样率。重要的是，这并不意味着每项EMA研究都应采用较高的采样率。在许多情况下，所研究的过程变化缓慢（例如，Courtet和Oli´e（2012年）指出，抑郁症患者的情绪存在昼夜波动，症状在早晨更严重；而Proudfoot等人（2014年）发现，双相情感障碍患者的情绪呈每周周期性变化）。奈奎斯特定理（Nyquist，1928年）为确定采样率提供了一条通用规则：采样率必须高于所研究现象频率的两倍（例如，如果某个过程以24小时为周期变化，研究人员应选择大约每12小时采样一次）。对于与事件相关的现象（例如物质使用），研究人员可以选择行为驱动时间表，这通常能达到合理的准确性；不过，这类时间表可能需要辅以预定设置，以捕捉未记录的事件。关于采样率对动力系统推断的影响的最新研究，参见Haslbeck和Ryan（2022年）。

2.2.2. 离散时间与连续时间

就分析而言，提示时间表会对建模方法的表现产生重大影响。如果提示之间的间隔大致相等（例如每天在大致相同的时间提示一次），研究人员可以使用方程 1–4 中描述的离散时间状态空间模型来建模动态。然而，如果提示时间表导致提示之间的间隔变化很大（例如采用行为驱动时间表，事件并非以相等的间隔发生，或测量出现聚集），研究人员则应使用连续时间状态空间模型。方程 1–4 中描述的模型的连续时间版本会将

方程 1 改写为随机微分方程：

$$\dot{\mathbf{x}_t} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t + \boldsymbol{\varepsilon}_t \tag{5}$$

其中，瞬时变化率 $\dot{\mathbf{x}_t}$ 是当前 x 值、 u 的任何扰动效应以及随机效应的函数。该模型可以视为一种隐马尔可夫模型，其中的状态是连续的，而非离散的。当连续时间模型更为合适时，使用离散时间模型可能会增加效应估计的偏倚（de Haan-Rietdijk、Voelkle、Keijsers 和 Hamaker，2017年），不过有研究表明，只要测量间隔的变异不大，就可以使用离散近似来建模生态瞬时评估数据（Boker、Deboeck、Edler 和 Keel，2010年；Tuerlinckx、Maris、Ratcliff 和 De Boeck，2001年）。

离散时间和连续时间状态空间模型的参数估计值在解释上有所不同；然而，这些差异源于模型的单位，而且在假设线性动力学的情况下，至少可以使离散时间和连续时间情形下的参数估计值相互等价。例如，若一个双状态离散时间状态空间模型的状态动态参数矩阵为 $\mathbf{A} = \begin{bmatrix} .5 & .2 \\ 0 & .5 \end{bmatrix}$，那么其对应的连续时间表示为 $\log\mathbf{A} = \begin{bmatrix} -.69 & .4 \\ 0 & -.69 \end{bmatrix}$。这两个参数矩阵的含义不同：$\mathbf{A}$ 表示从时间 t 到时间 $t+1$的状态间关系，而 $\log\mathbf{A}$ 表示时间 t的状态值与时间 t的一阶导数（即变化率）之间的关系。由于连续时间模型的动态参数是离散时间模型动态参数的指数化值，两者之间存在一一对应关系，因此可以认为这两种表示是等价的；但需要注意，这种等价关系仅在线性模型中成立。关于离散时间模型参数与连续时间模型参数之间关系的详细分析和讨论，参见 J. H. L. Oud（2007年）。

当然，离散时间与连续时间状态空间模型在解释上存在差异，但最大的区别在于估计方法，这也限制了可实现连续时间模型的 R 软件包数量。截至本文撰写时，dynr（Ou 等，2019年）、OpenMx（Boker 等，2023年）和 ctsem（Driver、Oud 和 Voelkle，2017年）是仅有的可实现连续时间状态空间模型的 R 软件包。如果研究人员怀疑模型中可能存在制度切换（见下文“平稳性”一节），或者拥有多被试数据，建议使用 dynr 软件包。如果研究人员关注混合效应连续时间模型，我们建议使用 ctsem。最后，在上述软件包中，OpenMx 是用途最广泛的潜变量建模软件包，研究人员可用它构建观测变量模型、潜变量模型、横断面模型和动态模型。关于连续时间状态空间模型的进一步介绍，可参阅 Lodewyckx、Tuerlinckx、Kuppens、Allen 和 Sheeber（2011年）；Oravecz、Tuerlinckx 和 Vandekerckhove（2011年）；J. Oud 和 Jansen（2000年）。

2.2.3. 突发次数

除了考虑提示将在何时以及如何发生之外，研究人员还必须考虑将发生多少组提示。许多研究通过将生态瞬时评估与源于决策点的其他研究设计相结合，采用单次突发设计或多次突发设计，从而发挥其优势。例如，多次突发方案是在治疗前设置一个生态瞬时评估数据收集间隔，然后在治疗后设置另一个独立的生态瞬时评估数据收集间隔——

。另一种选择是单次突发设计（如 MacKerron 和 Mourato（2013）所采用的设计）；当研究人员不太关注不同时间段之间的变化时，这种设计会很有用。总体而言，选择单次突发还是多次突发取决于研究问题：与提示时间表一样，在研究设计层面，应当基于理论谨慎作出决定，以便恰当地测量所研究的现象。有关单次突发设计与多次突发设计的示例，请参见 Bai、Elavsky、Kishida、Dvo˘r´akov´a 和 Greenberg（2020）；Fuller-Tyszkiewicz 等（2020）；Lay、Gerstorf、Scott、Pauly 和 Hoppmann（2017）；Malmberg、Lim、Tolvanen 和 Nurmi（2016）；Ram、Brinberg、Pincus 和 Conroy（2017）。

多次突发设计很容易分析，前提是突发内数据经过了恰当的分析，因为每次突发都可以单独分析，并在分析后进行比较。一个重要的考量是，研究中的过程可能会如何随着治疗进程或时间推移而变化。例如，在一项研究父母与青少年幸福感的研究中，Janssen 等人（2020年）采用了新冠疫情前和新冠疫情期间进行生态瞬时评估测量的多次突发设计；当代读者会明显看出这两个时期的过程存在差异。如果研究者认为这种差异在理论上有可能出现，那么就应优先采用多次突发设计。如果研究者认为纵向差异或前后测并不相关，那么单次突发就足够了。总体而言，数据收集设计中突发次数的设计，取决于研究者预期研究现象会发生何种变化。关于生态瞬时评估研究中单次突发与多次突发设计的进一步阅读，参见 Nesselroade（1991年）；Sliwinski（2008年）。

2.3. 时间趋势与夜间效应

到目前为止，我们一直隐含地假设，时间不会对所研究的过程产生任何系统性或周期性的趋势。然而，在对思想、感受和行为的研究中，几乎总是存在系统性趋势。例如，情绪中众所周知的“周末效应”（Stone, Schneider, & Harter, 2012），即周末的积极情感明显高于工作日。另一个与时间相关的趋势是皮质醇的昼夜节律。皮质醇是一种与应激反应相关的激素，其水平在醒来后立即达到最高值，随后通常会在一天之中逐渐下降（DeSantis et al., 2007）。更重要的是，人类具有周期性的生理节律，活动时段与睡眠时段交替出现，因此，任何EMA研究都必须考虑睡眠对其研究结果的影响，即使最终只是为了将这种影响排除为无关因素。

2.3.1. 时间趋势

图1中的面板B展示了两种时间趋势。蓝线表示随时间简单线性增长，而洋红线表示每个模拟周的“周末”期间出现增长。请注意，除了时间趋势之外，用于生成这些时间序列的其余参数均相同，这突显了时间趋势对时间序列轨迹的塑造作用。

状态空间建模可以通过多种方式来处理方程1–4中描述的两种时间趋势。时间可以编码为外生变量，并纳入$\mathbf{u}_t$中。根据编码方式的不同，这种外生变量方法可以表示多种类型的时间效应。假设存在如图1.B所示线性趋势的研究者，可以纳入研究开始以来的绝对时间（例如，$t = 1, 2, \ldots, T$）；而那些认

为理论上存在周末效应（或仅仅是特定时段的效应）的研究者，可以用0/1编码对其建模。最后，如果研究者认为存在更连续的波动或日内效应，可以通过编码时间来反映一天中的时刻或醒来后的时间，从而对其建模。不过，必须谨慎处理，尤其是在考虑多项研究的分析时。Shiffman、Stone 和 Hufford（2008年）建议，对时间的编码应能体现参与者时间线上的相对差异。例如，不使用一天中的绝对时间（如午夜以来的时间），而使用醒来后的时间。这样可以考虑到参与者日程安排的差异。关于生态瞬时评估数据收集中特定于时间或事件顺序的诸多问题，建议仔细阅读 S
hiffman 等人（2008年），以获取易于理解的概述。

2.3.2. 夜间效应

这里，我们用“夜间效应”来概括可能由睡眠导致的动态差异。例如，对某个人而言，情感过程在清醒时段可能相当稳定，但睡眠时段内不太可能出现同样稳定的动态（例如，情绪可能在睡眠期间重置到基线水平）。与时间趋势或不规则时间间隔相比，夜间效应较难建模，因为它们并非由所收集数据的时间安排差异造成，而是源于两次数据收集之间存在较长的时间间隔。此外，对许多参与者而言，这段时间（即夜间）在性质上与白天截然不同。

状态空间建模有三种方法可以处理夜间效应。首先，研究者可以使用连续时间状态空间方法，以应对两次数据收集之间的较长间隔。然而，使用连续时间模型处理夜间效应时包含一个重要假设：状态过程会在夜间逐渐回到其平稳分布。这意味着一天中最后一次观测与次日第一次观测之间的时间依赖性会减弱。其次，研究者在使用离散时间状态空间模型时，可以采用一种等效方法，即在夜间时段加入缺失观测，具体来说，加入数量足以近似覆盖整个夜间时段的缺失观测（如果白天 12 小时内有 5 次提示，那么夜间时段就加入 5 个缺失观测）。上述方法无法处理特定的夜间效应，例如情绪在睡眠一段时间后发生变化。要处理特定的夜间效应，第三种方法是像使用上述任一符合研究者需求的方法处理时间趋势那样，将时间编码为相对于夜间的时间。不过，针对特定夜间效应的外生变量设定需要结合具体现象加以考虑。例如，睡眠导致的情绪下降会在白天持续多久？这会影响研究者如何定义$\mathbf{u}_t$，因为它决定了研究者将时间变量的零点设在哪里。

一般而言，理解并考虑时间趋势和/或夜间效应，需要研究人员考量参与者的生活体验，以及他们的日常安排会如何影响行为结果。此外，必须谨慎地对时间进行编码，以确保估计效应在不同参与者之间具有可解释性。不幸的是，未能考虑时间趋势可能会导致估计参数出现严重偏倚（Shumway & Stoffer, 2011）。因此，无论时间趋势是与研究无关的干扰因素，还是具有科学研究价值，都应妥善处理。此外，这些时间趋势可能并非孤立发生，可以视为 $\mathbf{u}_t$ 的独立方面（即，时间趋势可以与.....同时建模

观察到的扰动）。由于大多数时间趋势都可以通过在扰动项$\mathbf{u}_t$中纳入时间来建模，因此大多数状态空间建模软件包都可以使用。当然，夜间效应的连续时间建模仅限于连续时间分析软件包，例如 ctsem 和 dynr。若想通俗易懂地了解这一重要主题，我们再次建议读者参阅 Shiffman 等人（2008年），以及 Chow、Zu、Shifren 和 Zhang（2011年）；Dethlefsen 和 Lundbye-Christensen（2006年）。

2.4. 缺失数据

缺失在生态瞬时评估数据收集中至关重要，因为该方案本身会给参与者带来很大负担，进而导致流失。研究人员必须调查缺失机制，才能继续开展研究，因为不同的缺失机制需要研究人员采取不同程度的干预。

2.4.1. 缺失数据机制

传统上，缺失机制分为完全随机缺失（MCAR：缺失不依赖于模型中的任何变量）、随机缺失（MAR：变量 Y 的缺失取决于模型中的另一个变量 X）和非随机缺失（MNAR：变量 Y 的缺失取决于 Y 本身的观测值）。例如，如果研究人员正在研究酒精使用情况，那么酒精使用情况报告的遗漏可能是由另一项研究变量（如就业状态）引起的（MAR）；在最糟糕的情况下，也可能是由酒精使用本身引起的（MNAR）。生态瞬时评估数据（EMA data）还带来了两个问题，使这一情况更加复杂。首先，数据在时间上具有相关性。其次，与此相关的是，时间序列中的缺失可能与时间有关，这正是时间依赖型随机缺失（TMAR：缺失取决于时间变量）和自回归时间依赖型随机缺失（ATMAR：结果变量的缺失与该变量先前的取值有关）两种情况下所出现的现象（Slipetz, Falk, & Henry, n.d.）。在前一种情况下，缺失通常表现出周期性，并且会出现多个连续的数据点缺失（例如，在一项饮酒研究中，我们可能会发现连续多天的夜间数据点均出现缺失）。目前已有多种方法可用于检验不同的缺失机制（Chen, Diao, & Qin, 2020; Jaeger, 2006; Little, 1988），但针对时间依赖型缺失机制（TMAR 和 ATMAR）的研究仍相对较少。

在大多数情况下，尤其是当数据属于随机缺失、依赖缺失值的随机缺失或非随机缺失时，生态瞬时评估数据中的缺失通常不可忽略（因为会造成较大的偏倚和变异性），需要明确加以处理。当数据属于完全随机缺失或时间依赖型随机缺失时，参数估计值的偏倚和变异性较小，因此问题没有那么严重。因此，必须特别谨慎地确定缺失机制。如果缺失数据处理不当，就会影响参数估计值的统计功效、效度和准确性。鉴于缺失数据在生态瞬时评估研究中十分常见，且会由此引发各种问题（尤其是在存在随机缺失、依赖缺失值的随机缺失或非随机缺失时），研究者最好在分析之前制定缺失数据计划。

2.4.2. 缺失数据的解决方案

通过研究设计可以尽量减少缺失，例如制定补偿安排、联系参与者，以及避免让参与者填写过长的调查问卷。尽管如此，仍有多种处理缺失数据的方法。单重插补和多重插

补方法都已得到应用，其中 MICE（van Buuren 与 Groothuis-Oudshoorn，2011年）和 Amelia（Honaker、King 与 Blackwell，2011年） R 软件包是目前开发最完善的多重插补系统。单重插补是指用估算值填补单个数据集中的缺失值；多重插补则是指用估算值填补原始数据集的多个版本，分别分析每个数据集，再汇总各数据集的结果。近期，这篇论文的作者发现，对于离散时间状态空间模型，采用多重插补处理题目层面数据的缺失并不能解决偏倚问题；而在观测缺失时，直接使用标准滤波方法（例如用于离散时间状态空间模型的卡尔曼滤波器）来插补状态值，在各种条件下的表现最佳（Slipetz 等人，未注明日期）。需要注意的是，经典卡尔曼滤波器仅适用于离散数据；不过，也有多种滤波方法能够以同样方式为连续时间和非线性状态空间模型插补状态值。

滤波方法是一组状态估计技术，可根据截至给定时间点的数据以及方程1-4中的模型参数，依次估计状态及其协方差。如前所述，经典卡尔曼滤波器适用于具有线性动力学的离散时间状态空间模型，但任何滤波技术都可视为一种状态值单重插补方法。虽然没有明确的规则规定缺失比例达到多少时滤波才能作为缺失数据校正工具有效运行，但作者发现，经典卡尔曼滤波器在缺失比例为30%时仍能成功恢复参数（Slipetz 等人，未注明日期）。对于离散时间模型中的缺失数据，我们建议使用 dlm 或 dynr 软件包，因为两者都能原生使用卡尔曼滤波器处理缺失数据。总的来说，在生态瞬时评估研究分析中处理缺失数据的方法是一个重要且亟待研究的领域。我们建议感兴趣的读者参阅我们的研究，以及 Padilla、Lagos-́Alvarez、Mateu 和 Porcu（2020年）、Pratama、Permanasari、Ardiyanto 和 Indrayani（2016年）以及 Walani 和 Cleland（2015）的研究，了解相关的初步探索。

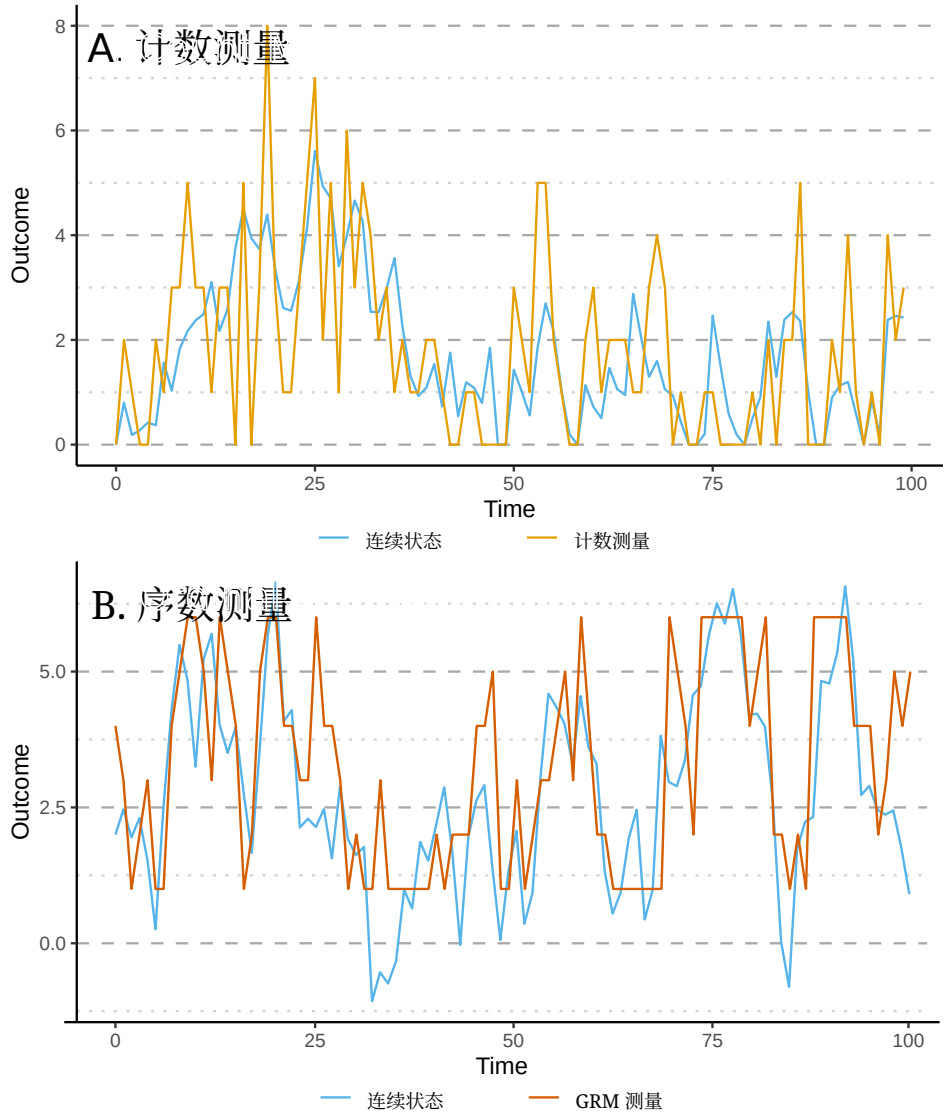

图 2. 图 A 展示了计数数据状态空间模型中的潜在状态轨迹（蓝线）及相应的计数变量（黄线），其中计数数据服从泊松分布。例如，对于单个状态空间模型，状态轨迹可以表示酒精使用的预期次数。图 B 展示了包含有序测量变量的状态空间模型，其中观测变量由分级反应模型（红线）建模，潜在状态轨迹以蓝线表示。该模型适用于李克特式测量变量。

2.5. 测量

状态空间模型包含一个测量模型，用于将未观测状态映射到测量所得的观测值。方程1-4中描述的模型意味着状态和观测值都是连续且服从正态分布的；可以将其理解为在某种程度上等同于潜变量模型中对潜变量和显变量指标具有多元正态性的假设（Bollen，1989）。然而，生态瞬时评估数据很少通过真正连续的观测值收集。相反，生态瞬时评估研究中的观测变量通常采用李克特量表（例如，1-5分，表示从同意到不同意）、事件指标（0/1，表示某事是否发生），或某种更为类别化的.....

分类测量。在这种情况下，计算均值并无意义，因为每个刻度之间的距离（例如，从1到2与从2到3）未知，因此均值将无法解释。从研究设计的角度来看，研究人员必须在参与者负担与问卷长度或每次提示时提出的问题类型之间取得平衡。对于常采用李克特量表的题项，一种提高测量精细度的潜在方法是使用滑块，让参与者在0到100之间的半连续量表上作答。遗憾的是，关于使用滑块代替李克特量表作答的证据并不一致。有些证据表明，滑块能够改善测量（Voutilainen, Pitk¨aaho, Kvist, & Vehvil¨ainen-Julkunen, 2016年）（因为年龄等混杂因素的影响较小，且天花板效应较弱）；但另一些证据则表明，滑块未必能改善测量（因为滑块的外观可能导致结果产生偏倚），还可能延长作答时间（Matejka, Glueck, Grossman, & Fitzmaurice, 2016），从而增加参与者负担，并加剧前文提到的流失问题。

2.5.1. 计数数据

虽然将李克特量表题项转换为滑块是一种选择，但对于生态瞬时评估研究中使用的许多题项类型而言，回答类别并非连续分布，即使在理论上也无法将其变为连续变量。以酒精使用相关题项为例。在大多数横断面研究中，酒精使用可能通过一种映射到分箱计数的有序测量进行测量。例如，一个常见的酒精使用题项可能是：“在过去两周内，您平均每天喝了多少杯酒”，回答选项为 0：不饮酒，1：1-3 杯，2：4-5 杯，以及 3：$5+$ 杯。这类题项乍看之下属于有序题项，因为每个回答类别都大于前一个类别，而回答类别之间的差异则按定义是不相等的。基于构念效度，使用这类数量/频率题项来测量物质使用一直受到广泛批评（McGinley 和 Curran，2014年）。对于使用有序分箱计数来测量物质使用等事件的研究者而言，一种替代方法是利用生态瞬时评估数据收集的实时特性，请参与者回忆自己当天的使用情况。这种做法产生的不是有序数据，而是计数数据（即当天饮用的酒杯数）。与使用有序分箱计数相比，这在测量方面是更好的做法，因为它避免了将多个计数范围合并在一起；然而，使用计数数据仍然需要改变状态空间模型中的测量部分。

从理论上讲，修改状态空间模型的测量部分，就是选择未观测状态变量与观测结果之间的正确关系。例如，如果认为酒精使用次数服从泊松分布，那么得到的（单一）状态空间模型将是

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t + \boldsymbol{\varepsilon}_t \tag{6}$$

$$\boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma}) \tag{7}$$

$$\mathbf{y}_t \sim \text{Pois}(x_t) \tag{8}$$

在这里，状态 x_t 被操作性定义为每日酒精使用的预期次数（因为只有一个观测结果）。此类系统的潜在状态轨迹及其对应的计数变量如图 2 的 A 面板所示。对于拥有多个状态变量或计数测量变量的研究人员，我们建议仔细地对状态与测量值之间的关系进行操作性定义，

也就是说，使用同一个状态估计值 x_t 来表示酒精使用和吸烟的平均次数，会忽略每日吸烟量通常远高于每日酒精使用量这一事实。如果研究者假设存在一个同时驱动酒精使用和吸烟的潜在状态，那么至少需要对状态空间模型进行参数化，以便对单一状态值进行正确缩放。例如，在状态为 x_t的单状态模型中，用 $Poisson(x_t)$表示酒精使用和吸烟是不正确的；但若加入缩放参数（即酒精使用 $\sim Poisson(\beta_1 x_t)$和吸烟 $\sim Poisson(\beta_2 x_t)$），就可以计算出适当的缩放值。最后，虽然我们没有明确讨论零膨胀分布的用法，但可以将分布中的零膨胀部分整合到状态空间模型中。

2.5.2. 有序、二分和分类测量值

对于自然具有顺序等级的测量变量（例如李克特量表），仍可指定适当的测量模型。我们为李克特量表变量推荐的一种测量模型是分级反应模型（Samejima，1969年），它将回答类别 k 或更高类别的累积概率建模为潜在状态的函数（在实践中，分级反应模型用于教育测验，因此这一潜在状态通常称为能力）。使用我们对测量变量和状态的状态空间建模记法，依赖于单个状态的单个条目的测量模型可表示为：

$$P(y_t > i) = \frac{1}{1 + \exp(-\alpha(x_t - \beta_i))} \tag{9}$$

其中， α 为区分参数，用于控制各类别的区分程度； β_i 为第 i个响应类别的阈值参数，表示响应高于 ith 个响应类别的概率为.5时所对应的临界点。观测到的分级反应模型过程及其底层状态的可视化示例见图2B。分级反应模型适用于李克特类型的响应，因为该模型假定较高的潜在状态值对应较高的响应类别。

与序数测量值的设定思路类似，可以通过使用适当的测量模型（即双参数逻辑模型，一种项目反应理论模型，其中正确作答的概率服从逻辑分布：Manapat & Edwards, 2022年），将二分测量指标整合到状态空间中。真正的类别变量作为测量值整合到状态空间模型中则更加困难，但可以通过**u** 矩阵轻松地将其整合为状态的外生预测变量。

2.5.3. 估计

尽管在状态空间建模框架下，完全可以针对有序、计数、二分和分类数据，从数学上提出非正态测量模型，但要估计这些模型却困难得多。我们列出的 R 软件包使用卡尔曼滤波；dse（Gilbert, P. D., 2006年或之后）和 FKF（Luethi, Erb, Otziger, McDonald, & Smith, 2022年）采用卡尔曼滤波或其众多扩展之一。这些滤波方法即使在最简单的传统卡尔曼滤波情况下，也依赖于状态和测量变量均服从正态分布这一假设。虽然扩展卡尔曼滤波和

无迹卡尔曼滤波是卡尔曼滤波的两种扩展（E. A. Wan & R. Van Der Merwe, 2000年; Simon J. Julier & Jeffrey K. Uhlmann, 1997年），可用于建模非线性动力学和状态—测量关系，但任何形式的卡尔曼滤波都不适用于非连续测量变量。少数将状态空间建模用于非连续测量变量的研究，要么将有序结果视为连续变量（van Rijn & Molenaar, 2005），要么采用贝叶斯估计方法（de Haan-Rietdijk, Kuppens, et al., 2017年; DeYoreo & Kottas, 2017年; Imai, Jain, & Ching, 2009），而这需要编写定制建模代码来指定状态空间模型。根据我们对现有工具的审查，目前没有可直接用于拟合非连续测量值状态空间模型的软件包，只有一个例外：截至本文撰写时，本文作者已开发出一个 R 软件包，可在频率学派框架下简单实现分级反应模型测量状态空间模型，适用于生态瞬时评估（EMA）数据中常见的李克特式反应时间序列。该软件包 genss 公开提供（Falk, Slipetz, & Henry, 2023年），未来还将扩展纳入其他测量模型。

2.6. 扰动与误差

生态瞬时评估数据收集本质上处于自然情境且具有时间性，即在给定时间区间内收集个人现实世界体验的数据。个体会经历任意数量的外部事件，这些事件会影响所研究的现象。虽然我们可以认为这些事件产生的影响服从随机分布，并由许多所使用模型中包含的白噪声误差项表示，但当然，参与者生活中的某些事件会对所研究的过程产生系统性影响。例如，在研究情感过程时，情感的日常扰动可能总体上没有影响，但重大的生活事件（例如丧亲）会产生系统性影响；除非对其进行建模，否则这些影响将使最终模型产生偏倚。

请考虑方程1—4中每个“误差”项的具体含义。状态转移的创新项 ε_t表示潜在状态过程中的波动。在生态瞬时评估的情境下，该项包含生活事件对状态变量的集体影响，并作出了一个强假设，即生活事件的总体影响约为0。测量方程的误差项 ν_t 是测量误差，即测量设备不准确的程度；这里同样作出了一个强假设，即这种测量误差平均起来不会产生影响。第一个假设在生态瞬时评估数据收集的情境下问题最大，因为它假设没有任何生活事件能够引起状态过程的永久性变化，只能引起最终会消退的变化。

我们将那些会导致状态（因此也会导致测量值）发生持久或重大变化的事件称为重大扰动。工程问题中采用的状态空间建模技术有不少可用于检测和消除扰动或异常，但绝大多数技术很难应用于生态瞬时评估数据，甚至可能根本无法应用。这是因为这些方法需要准确估计系统动力学，以及测量过程和状态过程中的预期误差；同时，扰动相对于这些模型必须是真正的离群值。在生态瞬时评估数据收集与分析中，我们面临双重问题：正确的模型尚不明确，因此需要对其进行估计；而且可能对参与者生活造成系统性影响的重大扰动不计其数。因此，在生态瞬时评估数据建模中，解决重大扰动问题的办法不是尝试自动检测并移除它们，而是利用这种收集方式，收集有关日常生活事件的细致信息。

2.6.1. 检测和建模重大扰动

我们建议每天增加一个问题，该问题适用于大多数（甚至所有）生态瞬时评估研究：“今天有没有发生什么令人难忘的事，无论是好事还是坏事？”如果参与者对这一条目作出肯定回答，则研究方案应要求其简要描述所发生的事情。这个问题可以根据具体结果和收集时间范围进行调整，并且只需提供可能存在重大扰动的观测数据。MacKerron 和 Mourato（2013）提供了一个例子。他们关注的是幸福感与环境之间的关联，同时也收集了参与者与谁在一起以及当时在做什么的信息。关键在于通过自由回答收集扰动信息，然后将其量化编码为不同类别，以便明确纳入状态动力学方程（公式 1），具体来说，是纳入$\mathbf{u}_t$ 扰动项。不过，在指定扰动如何 影响状态动力学时，必须谨慎。

例如，简单地将某个生活事件编码为仅影响一个时间点（即，对单个 t使用$\mathbf{u}_t = 1$）不会导致该生活事件产生任何持久影响；而将某个重大生活事件（即，在一项抑郁症研究中发生丧亲）编码为事件发生后所有 t 均使用$\mathbf{u}_t = 1$ ，则意味着这种扰动会产生持久影响。这两种选项之间的折中做法，是将事件编码为一个递减序列（即，事件发生后 t 使用$\mathbf{u}_t = 1, .5, .25, .125, \dots$）。如何在这些模型中纳入扰动，首先必须依据研究者对扰动可能如何影响所关注结果的理论理解来决定。我们建议，在几种编码选项之间进行评估时采用实证方法，并根据不同的扰动设定，使用 AIC 或 BIC 等信息准则考察模型拟合优度。需要牢记的是，尽管并非所有模型都会被采用，研究者仍应报告所有分析的结果，以便读者评估结果对模型设定的敏感性。多宇宙分析（Steegen、Tuerlinckx、Gelman 和 Vanpaemel，2016年）也适用于此处。

幸运的是，从这些模型的实现角度来看，纳入扰动项很简单。前文提到的用于拟合简单和更复杂状态空间模型的软件包 dse、ctsem 和 dynr 都可以纳入扰动项及其他外生预测变量。关于不同类型扰动和干预的具体设定，建议读者参阅 Ryan 和 Hamaker（2022年）以及 Fr¨uhwirth-Schnatter 和 Wagner（2010年）。

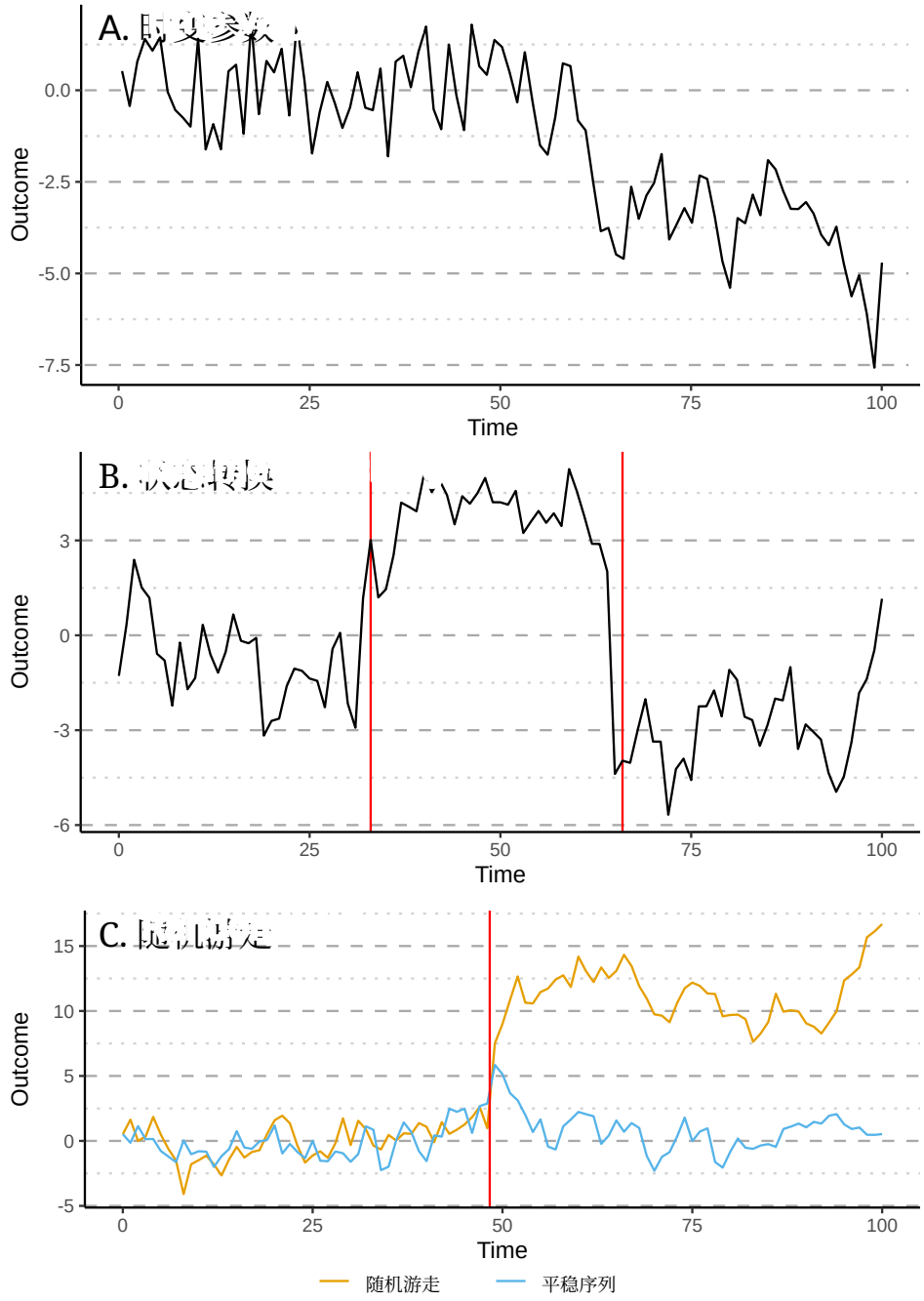

图3. 该图展示了三种非平稳过程。面板A是一个自回归过程的示例，其参数随时间变化（自回归参数在 $t = 0$ 时从 0 开始，以 S 形方式增加，到 $t = 100$ 时达到 1）。面板B是一个状态转换过程的示例（均值在 $t = 33$ 时从 0 变为 3，并在 $t = 66$ 时再次从 3 变为 -3）。面板C是一个单变量随机游走过程（黄色线）和一个平稳过程（蓝色线）的示例，两者的误差方差均为 ε，且在 $t = 50$ 时受到扰动。对于平稳过程，扰动导致数值上升，随后又回到围绕均值 0 波动的状态。对于随机游走过程，扰动导致数值上升，而上升后不会回落。

2.7. 平稳性

应用于时间序列模型的平稳性定义较为复杂，包含许多技术细节，但大致可以概括如下：如果数据生成过程在每个时间点都相同，则该时间序列是平稳的。虽然平稳性是一项相当严格的性质，但时间序列可能存在多种非平稳的情况。我们将讨论三种最可能适用于心理学数据的非平稳性类型：时变参数、制度切换和随机游走。

2.7.1. 时变参数

状态空间框架中的时变参数允许控制状态动力学或测量特性的参数随时间变化。这涵盖了多种情况，从状态动力学**A** 随时间缓慢变化（例如，由于实施了某项干预），到测量的含义突然改变，进而导致**H** 矩阵发生变化。图3的A面板展示了一个自回归过程的示例，其中参数随时间变化。在该例中，自回归参数在 $t = 0$时从0开始，以S形方式递增，并在 $t = 100$时达到1。这使得序列开头的时间点在时间上实际上互不相关，而序列末尾的时间点则在时间上高度相关。Z. F. Fisher等人（2022年）提供了一个可能具有时变参数的心理过程示例：参与者参加正念冥想工作坊之前、期间和之后的正性与负性情感。利用这些数据，Fisher及其同事发现，在工作坊期间，正性和负性情感的自回归参数均逐渐减弱。这反过来表明，对某些参与者而言，正念冥想练习可能会导致情感状态在时间上脱钩，使正性和负性情感都较少依赖先前的取值。时变参数状态空间模型可作为有用的工具，帮助研究者探索干预或生活事件对心理动态的影响；然而，读者必须了解其实施过程中涉及的若干问题。

首先，截至撰写本文时，目前尚无通用时变参数状态空间模型的开源 R 实现，先前的应用也都使用了定制实现。其次，指定时变参数模型时，需要仔细考虑哪些参数可以随时间变化，以及它们如何变化；因为除少数特殊情况外，无法指定一个让每个参数都任意随时间变化的状态空间模型。通用时变参数状态空间模型的开发与实现是方法学研究的活跃领域，我们预计未来几年将会开发出更多易于获取的开源实现。不过，有一种特定类型的时变参数状态空间模型已有易于获取和使用的实现：制度切换状态空间模型。

2.7.2. 制度切换

制度切换状态空间模型是一种特殊类型的时变状态空间模型，允许动态或均值水平发生突然变化，而这些变化在给定的时间区间（也称为“状态”）内保持一致（Kim & Nelson, 1999）。通常，在一个状态内，状态空间模型是平稳的，而在不同状态之间，参数则可能发生变化。制度切换动态可能发生在心理时间中。

由于种种原因，时间序列可能会发生变化。例如，Chow 等人（2011年）使用制度切换微分方程模型，表明母子二元组中儿童的行为会出现动态上的突然转变，从试图待在母亲身边转变为更具探索性的行为。图3B面板展示了制度切换。该时间序列在 $t = 33$ 处均值从0变为3，并在$t = 66$ 处从3变为-3。无论是在离散时间还是连续时间中，制度切换状态空间模型（或制度切换观测时间序列模型）的设定和拟合都格外简单，这得益于 Ou 及其同事在开发 dynr 方面所做的大量工作（Ou 等人，2019年）。如果数据存在若干稳定时期，但这些时期之间存在差异，研究者就应使用制度切换模型（例如上述比较新冠疫情前和新冠疫情期间的情形）。

2.7.3. 随机游走过程

随机游走非平稳性是本文讨论的非平稳性形式中，唯一并非严格意义上的时变参数模型示例。简单来说，随机游走过程是指无论变化源于系统性扰动还是常规波动，变化都会永久保留的过程。数学上，单变量随机游走过程可表示为 $x_{t+1} = x_t + \varepsilon_t$，而多变量随机游走过程的设定则更为复杂。图3C面板展示了误差方差相同、均为ε的单变量随机游走过程（红线）和平稳过程（蓝线）。在 $t = 50$ ，引入了一次扰动，使序列的值增加一个固定常数。请注意随机游走过程与平稳过程之间的差异：平稳过程在 $t = 50$ 有所上升，但随后很快回到围绕模型隐含均值 0 波动的状态。然而，随机游走过程因扰动而上升，并保持在较高水平，不会回到先前的水平。

平稳时间序列与随机游走时间序列之间的差异，对心理过程的建模有诸多影响。最重要的是，正如前文在扰动一节中所提到的，如果底层模型是平稳的，就必须将干预、治疗或扰动的效应设定为持续存在；否则，干预的效应会逐渐减弱并消失。未考虑平稳性的影响而错误设定干预效应，会导致效应估计减弱、检验效能降低。当研究动态中发生的变化是永久性的时，研究者应使用随机游走过程；也就是说，情绪的周末波动采用制度切换就足够了，而重大生活事件（如亲人离世）则最好使用随机游走过程。拟合随机游走状态空间模型，可以通过将自回归参数约束为 1 来实现（对于单状态模型而言）。对于多个状态，拟合随机游走模型更加困难，因为自回归效应与交叉回归效应必须保持平衡。对于单状态模型，可以使用任何允许固定参数的状态空间建模软件包。截至本文撰写时，尚无可用的多状态随机游走状态空间模型实现。

3. 总结

生态瞬时评估研究有潜力检验和探索心理现象中的广泛时间趋势。为此，一种有前景的分析框架是状态空间建模。使用这些模型时，需要考虑以下若干事项

表 1. 问题与建议1

ssue	建议	R 软件包
个体化与普遍化建模		
我没有理论依据认为该现象的 动态在所有个体之间都相同。	采用个体化方法（N = 1）	dynr；dlm； gimme；genss
该现象部分具有个体化特征，而我的模型 包含潜变量，或者我关注亚组。	使用 GIMME	gi mme
这种现象在一定程度上具有个体化特征，而且我有一个 参数数量较多的观测变量模型。	使用正则化的部分个体特异方法 例如 multivar。	multivar
这一现象具有一定的个体化特征，但我没有 很多变量。	使用多层向量自回归来允许 关系存在个体间差异	mlVAR
数据收集时间表		
我正在研究的现象以.....为尺度变化。 X 小时/天/周。	每隔 X*2 小时/天/周收集一次回答。	Any
这些提示大致间隔均匀。	使用离散时间模型	dynr；dlm；OpenMx
各次提示之间的间隔差异很大。	使用连续时间模型	dynr；ctsem；OpenMx
我不关注治疗前后的变化， 或者我没有在研究中期进行治疗。	采用单次突发设计	Any
我担心研究过程中出现的动态变化 （由治疗/干预导致）	采用多次突发设计	Any
时间趋势与夜间效应		
我认为我的数据存在时间趋势。	将时间编码为外生变量，并将其纳入 u_t （编码选项见正文）	Any
我认为这种现象发生了显著变化 在夜间时段（即睡眠期间负性情感降低）	将时间编码为外生变量 并将其纳入 u_t	ctsem；dynr
缺失数据		
我的数据属于随机缺失、非随机缺失或依赖缺失值的随机缺 失。	刻画并描述缺失问题，并 使用缺失数据方法调整模型/数据。	Any
我需要一种解决不可忽略的缺失问题的方法， 适用于离散时间模型。	标准滤波技术能够很好地处理 缺失。与成熟的 缺失数据方法（如多重插补）进行比较。	dynr；dlm
测量		
我有计数数据（例如物质使用次数、行为次数）	为以下内容选择适当的分布 计数变量（例如，泊松分布、负二项分布）	不可用
我有序数数据（例如，李克特量表回答）。	使用有序测量模型 例如分级反应模型。	genss
我有二分类数据（例如，症状是否存在）	使用逻辑回归/Probit 测量模型。	不可用
扰动		
本研究没有特定的事件/扰动 明确有兴趣研究。	询问：“有没有什么值得记住的积极方面 或负面，今天会发生吗？”	N/A
有些扰动并未产生持久影响。 关于这些现象。	将代码表示为单个时间点（u-t = 1）或短递减序列 （例如u t[+1,2,3] = [1,.5,.25]）	dse；ctsem；dynr
存在一种影响持久的重大扰动， 例如治疗或干预。	对所有时间点均保持不变的代码 发生后（例如 $u_t = 1$ 用于 $t >= X$）	dse；ctsem；dynr
存在一种影响会逐渐消退的扰动	将其编码为递减序列 （例如 $u_t[+1, 2, 3, \dots] = [1, .5, .25, \dots]$）	dse；ctsem；dynr
我如何知道自己纳入了正确的扰动?	比较包含和不包含 扰动的模型，使用赤池信息准则或贝叶斯信息 准则	dse；ctsem；dynr
平稳性		
我认为状态动力学正在随时间变化。	使用时变参数模型	不可用
我认为动态或均值会突然切换 这些状态会在一段时间内保持稳定，随后再次发生切换。	使用机制转换模型	dynr
我认为动态或均值发生了永久性变化。	使用随机游走模型	Any

这些因素必须在设计和分析中加以考虑。本文列举了其中的一些考量，并提出相应建议，以帮助成功开展生态瞬时评估研究的设计与分析（建议概要见表 1）。在规划研究时，这些考量的必要程度各不相同。

3.1. 研究设计

生态瞬时评估研究应旨在捕捉研究现象动态的个体间差异。因此，当理论原因不允许采用普遍化模型时，我们建议使用个体化或部分个体化建模，因为假设同一过程适用于所有研究参与者的普遍化模型，可能会掩盖相关的个体动态。此外，建模数据时必须考虑时间趋势，并特别注意对夜间和周末的影响进行建模，或证明研究变量不存在相关的时间趋势。研究者必须确保提示时间表能够充分追踪现象的动态（即，日内快速变化的过程比日内相对稳定、但在数周内发生变化的过程需要更多测量值），同时兼顾参与者负担。我们建议研究者优先提升测量质量（例如，增加每个构念的条目数），并采用有助于提高参与者响应率的做法，而非使用可能在先验上有更高响应率的简短问卷。我们认识到，在测量与负担之间取得平衡的最佳生态瞬时评估研究设计高度依赖具体主题，因此鼓励开展关于最佳研究设计的研究，以明确最佳实践。此外，生态瞬时评估研究者应制定缺失数据处理计划。近期研究表明，状态空间模型能够较好地应对缺失数据，因为它们可以利用先前时间点的信息，估算缺失指标时间点的状态（Slipetz 等人，未注明日期），卡尔曼滤波器便是如此；然而，这仍是方法学研究中尚未解决的问题，我们鼓励研究者在分析中描述缺失模式，并明确说明所采用的缺失处理方法。

3.2. 模型设定

关于状态空间模型的设定，研究者应考虑使用连续时间模型，尤其是对于那些测量一天中快速波动的构念（如情感）的设计。此前针对时间点间隔不等的情况下离散时间模型与连续时间模型进行的比较表明，离散时间模型能够稳健地应对时间间隔中少量的变异性（Loossens、Tuerlinckx 和 Verdonck，2021），而连续时间模型也可以离散化，从而使其能够与离散时间模型进行比较。即便如此，在解释连续时间模型的参数时仍须谨慎，因为其单位为瞬时变化率的效应，无法直接与使用离散时间模型计算得出的效应进行比较（J. Oud 和 Jansen，2000）。对于生态瞬时评估（EMA）研究者而言，指定正确的测量模型同样是一项值得关注的问题。然而，缺乏能够灵活实现不同测量结构的工具，限制了正确设定模型的使用。在开发出一种能够拟合状态空间模型中非连续测量模型的通用解决方案之前，EMA研究者应考虑使用题目的准连续版本（即0–100滑动量表），而不是选项较少的序数尺度。我们提出这一建议时仍持审慎态度，因为更多的方法学研究……

在评估滑动量表能否有效捕捉反应中的个体差异方面，仍需开展实证研究（即反应量表的精细程度究竟达到何种程度才算过于精细）。使用非连续结果数据的研究者应注意，将缺乏适当的测量模型视为一项局限性。最后，EMA研究者应考虑使用非平稳模型，但前提是有充分的理论依据支持这一做法。之所以需要谨慎，是因为非平稳模型仅由于其复杂性更高，就更容易导致数据过拟合，因此需要对其结果进行审慎解读。

3.3. 局限性与未来方向

出于简洁性考虑以及技术限制，本文省略了部分考量因素和建议。例如，我们侧重于定量测量（如序数测量），而非需要进行编码处理的定性测量（如开放式问题）。感兴趣的读者可参阅 Mayring（2007年）和 Britt（1997年）的相关研究。我们还假设所研究的潜在状态过程具有线性动力学。实际上，完全可以构建状态动力学为非线性、状态误差分布不服从正态分布，甚至状态本身为非连续的状态空间模型。有关非线性状态空间模型的描述，请参阅 Priestley（1988年）。在频率学派框架下对非连续测量变量进行建模，是当前研究的前沿领域。尽管作者已借助 genss 软件包迈出了初步步伐，但仍有许多工作有待开展。我们确定了以下几个亟待推进的未来研究方向：

- **非正态测量模型** -目前可用的状态空间建模实现无法整合非正态测量模型。这是心理学研究者亟需解决的问题，因为测量模型设定错误可能会对推断产生巨大影响。
- **心理学/行为科学中状态空间模型的数据需求** - 由于大多数关于状态空间模型的研究都是在工程学框架下开展的，因此，针对这些模型在心理科学典型情境下的性能评估，相关方法学研究仍然较少。关于效应量、时间点数量以及参与者数量如何影响统计功效的研究，将为研究设计提供极大的帮助。
- **多参与者状态空间模型**- 虽然已有一些研究使用观测变量进行多参与者时间序列分析，但鲜有研究探讨如何将状态空间建模框架应用于多个参与者的数据。多参与者时间序列分析方面的进展将很有帮助，因为并非所有研究都完全采用个体化方法。
- **用于心理/行为数据的复杂系统模型** ——心理学中一个新兴的思路是，心理系统属于复杂系统。然而，我们用于多变量时间序列（即具有线性动力学的时间序列）的标准模型无法捕捉任何复杂系统行为。非线性模型能够表现复杂系统行为，但需要进行谨慎的设定与构建。我们需要从心理学理论和定量发展两个角度，进一步研究“心理系统是复杂系统”这一命题的确切性质。未来的研究方向包括非平稳/非线性状态空间模型，以及呼吁开展更多理论工作，将心理学理论映射到具体的数据分析模型。

3.4. 结论

状态空间建模框架为时间序列数据，尤其是生态瞬时评估（EMA）数据的建模，提供了一条富有成效的途径。本文涵盖的主题多样性展示了该方法的灵活性。研究者决策点的多样化也意味着，研究者需要在研究设计和建模决策中保持深思熟虑和严谨审慎。总的来说，本文旨在支持EMA研究者做出这些选择。

参考文献

Bai, S., Elavsky, S., Kishida, M., Dvoˇr´akov´a, K., & Greenberg, M. T.（2020年）。正念训练对大学生日常压力反应的影响：一项随机对照试验的生态瞬时评估。《正念》, *11*(6), 1433–1445。Boker, S. M., Deboeck, P. R., Edler, C., & Keel, P. K.（2010年）。时间序列导数的广义局部线性近似。载于人类动力学建模的统计方法：跨学科对话。（第161–178页）。美国纽约州纽约：Routledge/Taylor & Francis Group。Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., . . . Niesen, J.（2023年）。OpenMx 2.21.1 用户指南。Bollen, K. A.（1989年）。含潜变量的结构方程。英格兰牛津：John Wiley & Sons。（页数：xiv, 514）Britt, D. W.（1997年）。建模的概念性介绍：定性与定量视角。Chen, X., Diao, G., & Qin, J.（2020年）。非忽略缺失数据问题中缺失机制函数的基于伪似然的估计与检验。《斯堪的纳维亚统计学杂志》, *47*(4), 1377–1400。（出版商：John Wiley & Sons, Ltd）Chow, S.-M., Zu, J., Shifren, K., & Zhang, G.（2011年）。时变参数动态因子分析模型。《多变量行为研究》, *46*(2), 303–339。Courtet, P., & Oli´e, E.（2012年）。昼夜节律维度与抑郁症严重程度。《欧洲神经精神药理学》, *22*, S476–S481。de Haan-Rietdijk, S., Kuppens, P., Bergeman, C. S., Sheeber, L. B., Allen, N. B., & Hamaker, E. L.（2017年）。关于混合马尔可夫模型在密集纵向数据中的应用。《多变量行为研究》, *52*(6), 747–767。（出版商：Routledge）de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L.（2017年）。对间隔不等的经验抽样法数据进行离散时间与连续时间建模。《心理学前沿》, *8*, 1849。DeSantis, A. S., Adam, E. K., Doane, L. D., Mineka, S., Zinbarg, R. E., & Craske, M. G.（2007年）。社区青少年样本中皮质醇昼夜节律的种族/族裔差异。《青少年健康杂志》, *41*(1), 3–13。Dethlefsen, C., & Lundbye-Christensen, S.（2006年）。在 R 中构建状态空间模型：重点关注纵向回归模型。《统计软件杂志》, *16*(1), 1–15。（栏目：文章）DeYoreo, M., & Kottas, A.（2017年）。用于非平稳时间序列的贝叶斯非参数马尔可夫模型。《统计与计算》, *27*(6), 1525–1538。Ditzen, B., Hoppmann, C., & Klumb, P.（2008年）。积极的伴侣互动与日常皮质醇：亲密关系的压力保护作用。《身心医学》, *70*(8), 883–889。Driver, C. C., Oud, J. H. L., & Voelkle, M. C.（2017年）。使用 R 软件包 ctsem 进行连续时间结构方程建模。《统计软件杂志》, *77*(5), 1–35。E. A. Wan, & R. Van Der Merwe。（2000年）。用于非线性估计的无迹卡尔曼滤波器。载于*IEEE 2000*年信号处理、通信与控制自适应系统研讨会论文集（目录号*00EX373*）（第153–158页）。（期刊缩写：Pro-

信号处理、通信与控制自适应系统研讨会论文集（目录编号：00EX373） Epskamp, S., Deserno, M. K., & Bringmann, L. F. (2021). mlvar：多层向量自回归。（R 软件包版本 0.5） Falk, A., Slipetz, L., & Henry, T. (2023年5月). netlabuva/genss：genss v0.1.0——初始预发布版本。获取自 https://doi.org/10.5281/zenodo.7887019 Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). 群体到个体的可推广性缺失对人类受试者研究构成威胁。美国国家科学院院刊，*115*(27)，E6106–E6115。（出版方：美国国家科学院院刊） Fisher, Z. (2022). multivar：多主体向量自回归（multi-var）模型的惩罚估计。（R 软件包版本 1.1.0） Fisher, Z. F., Chow, S.-M., Molenaar, P. C. M., Fredrickson, B. L., Pipiras, V., & Gates, K. M. (2022). 一种用于估计平滑时变参数的平方根二阶扩展卡尔曼滤波方法。《多变量行为研究》，*57*(1)，134–152。 Foster, K. T., & Beltz, A. M. (2018). 推进成瘾行为研究中移动评估数据的统计分析：三种以个体为导向的技术入门。《成瘾行为》，*83*，25–34。 Fried, E. I., & Nesse, R. M. (2015). 抑郁症并非一种一致的综合征：对 STAR*D 研究中独特症状模式的调查。情感障碍杂志，*172*，96–102。 Fr¨uhwirth-Schnatter, S., & Wagner, H. (2010). 针对高斯和部分非高斯状态空间模型的随机模型设定搜索。计量经济学杂志，*154*(1)，85–100。 Fuller-Tyszkiewicz, M., Krug, I., Smyth, J. M., Fernandez-Aranda, F., Treasure, J., Linardon, J., . . . Shatte, A. (2020). 饮食失调症状严重程度的状态标记。临床医学杂志，*9*(6)，1948。 Gates, K. M., & Molenaar, P. C. (2012). 群体搜索算法能够在同质与异质样本中恢复个体的有效连接图。神经影像，*63*(1)，310–319。 Gelman, A. (2006). 多层（分层）建模：它能做什么，不能做什么。《技术计量学》，*48*(3)，432–435。（出版方：泰勒与弗朗西斯出版集团） Gilbert, P. D. (2006年或之后). 简明用户指南：动态系统估计。 Haslbeck, J. M. B., & Ryan, O. (2022). 从心理学时间序列中恢复个体内动态。《多变量行为研究》，*57*(5)，735–766。 Henry, T., Feczko, E., Cordova, M., Earl, E., Williams, S., Nigg, J., . . . Gates, K. (2018). 使用验证性子组划分GIMME比较不同群体之间的有向功能连接。神经影像，*188*。 Honaker, J., King, G., & Blackwell, M. (2011). Amelia II：一个用于处理缺失数据的程序。《统计软件杂志》，*45*(7)，1–47。（栏目：文章） Imai, S., Jain, N., & Ching, A. (2009). 动态离散选择模型的贝叶斯估计。*Econometrica*，*77*(6)，1865–1899。（出版方：John Wiley & Sons, Ltd） Jaeger, M. (2006). 关于随机缺失假设的检验。载于 J. F¨urnkranz, T. Scheffer, & M. Spiliopoulou（编），机器学习：*ECML 2006*（第671–678页）。柏林、海德堡：施普林格·柏林·海德堡。 Janssen, L. H. C., Kullberg, M.-L. J., Verkuil, B., van Zwieten, N., Wever, M. C. M., van Houtum, L. A. E. M., . . . Elzinga, B. M. (2020). 新冠疫情是否影响父母与青少年的幸福感？一项关于每日情感与养育的生态瞬时评估研究。*PLOSONE*，*15*(10)，e0240962。 Kalman, R. E. (1960). 线性滤波与预测问题的一种新方法。基础工程学杂志，*82*(1)，35–45。 Kim, C.-j., & Nelson, C. R. (1999). 具有机制转换的状态空间模型：经典与

吉布斯抽样方法及其应用. 马萨诸塞州剑桥，英格兰伦敦：麻省理工学院出版社。Lane, S., Gates, K., Fisher, Z., Arizmendi, C., Molenaar, P., Hallquist, M., . . . Alvarez, S. C. (2022). gimme：群组迭代多模型估计。（R 软件包版本 0.7-8）Lane, S., Gates, K., Pike, H., Beltz, A., & Wright, A. (2018). 揭示流动评估数据中的一般、共享和独特时间模式。《心理学方法》,*24*。Lay, J. C., Gerstorf, D., Scott, S. B., Pauly, T., & Hoppmann, C. A. (2017). 神经质和外向性会放大回顾性与同步情感报告之间的差异。《人格杂志》, *85*(6), 817–829。Levinson, C. A., Williams, B. M., Christian, C., Hunt, R. A., Keshishian, A. C., Brosof, L. C., . . . Ralph-Nearman, C. (2023). 使用个体化模型为进食障碍治疗提供个性化方案：一项开放式系列试验。《咨询与临床心理学杂志》, *91*, 14–28.（地点：美国；出版商：美国心理学会）Little, R. J. A. (1988). 对多变量数据中完全随机缺失的检验。《美国统计学会杂志》, *83*(404), 1198-1202。Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N. B., & Sheeber, L. (2011). 情感动力学的层级状态空间方法。层级贝叶斯模型专题,*55*(1), 68–83。Loossens, T., Tuerlinckx, F., & Verdonck, S. (2021). 根据预测准确度比较情感过程的连续时间与离散时间建模。《科学报告》,*11*(1), 6218。Luethi, D., Erb, P., Otziger, S., McDonald, D., & Smith, P. (2022). Fkf：快速卡尔曼滤波器。（R 软件包版本 0.2.4）MacKerron, G., & Mourato, S. (2013). 自然环境中的幸福感更高。《全球环境变化》, *23*(5), 992–1000。Malmberg, L.-E., Lim, W. H. T., Tolvanen, A., & Nurmi, J.-E. (2016). 学习体验中的学生内变异，以及教师对学生任务专注度的感知。《前沿学习研究》, *4*(5), 62–82。Manapat, P. D., & Edwards, M. C. (2022). 检验等级反应模型和二参数逻辑斯蒂模型在构念正态性违背情况下的稳健性。《教育与心理测量》, *82*(5), 967-988。Matejka, J., Glueck, M., Grossman, T., & Fitzmaurice, G. (2016). 视觉外观对连续滑块和视觉模拟量表表现的影响。载于*2016*年计算机系统中的人为因素*CHI*会议论文集 (第 5421–5432 页)。美国加利福尼亚州圣何塞：美国计算机协会。Mayring, P. (2007). 混合定性与定量方法。载于（第 27–36 页）。荷兰莱顿：Brill。McDevitt-Murphy, M. E., Luciano, M. T., & Zakarian, R. J. (2018). 在成人治疗中使用生态瞬时评估与干预。《焦点：精神病学终身学习杂志》, *16*(4), 370–375。McGinley, J. S., & Curran, P. J. (2014). 将分箱计数定义的有序条目相乘所引发的效度问题：酒精使用数量-频率测量的应用。《方法论》, *10*(3), 108–116。McKee, K., Russell, M., Mennis, J., Mason, M., & Neale, M. (2020). 情绪调节动态可预测高风险青少年的物质使用。《成瘾行为》, *106*, 106374。Molenaar, P. C. M. (2004, 10月). 关于作为个体化科学的心理学宣言：让个体回归科学心理学，这一次永不离开。《测量：跨学科研究与观点》, *2*(4), 201–218。（出版商：Routledge）Molenaar, P. C. M., & Beltz, A. M. (2020). 个体建模：连接规律研究与个体研究的分析层次。载于 A. G. C. Wright & M. N. Hallquist（编），《剑桥临床心理学研究方法手册》（第 1 版，第 327–336 页）。剑桥大学出版社。

Mote, J., & Fulford, D. (2020). 精神分裂症患者日常社会体验的生态瞬时评估：一项系统综述。《精神分裂症研究》, *216*, 56–68。Nesselroade, J. (1991). 发展织体的经纬。Nyquist, H. (1928). 电报传输理论中的若干主题。《美国电气工程师学会汇刊》, *47*(2), 617–644。Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). 用于情感动力学的分层潜在随机微分方程模型。《心理学方法》, *16*(4), 468–490。Ou, L., Hunter, M. D., & Chow, S.-M. (2019). dynr 有什么用：一个用于 R 中线性和非线性动态建模的软件包。《*R*期刊》, *11*, 1-20。Oud, J., & Jansen, R. (2000). 使用结构方程模型对面板数据进行连续时间状态空间建模。《心理计量学》, *65*, 199–215。Oud, J. H. L. (2007). 交叉滞后面板设计中互惠关系的连续时间建模。载于《动力系统的数据分析技术》。(第 87–129 页)。美国新泽西州马瓦：Lawrence Erlbaum Associates出版社。Padilla, L., Lagos-́Alvarez, B., Mateu, J., & Porcu, E. (2020). 基于卡尔曼滤波，对含缺失数据进行时空自回归估计与预测。《环境计量学》,*31*(7), e2627。（出版商：John Wiley & Sons, Ltd）Petris, G. (2010). 用于动态线性模型的 R 软件包。《统计软件杂志》,*36*(12), 1 – 16。（栏目：文章）Pratama, I., Permanasari, A., Ardiyanto, I., & Indrayani, R. (2016). 时间序列数据缺失值处理方法综述。（页数：6）Priestley, M. B. M. B. (1988). 非线性与非平稳时间序列分析 / *M.B. Priestley*。圣迭戈：伦敦。Proudfoot, J., Whitton, A. E., Parker, G., Manicavasagar, V., Nicholas, J., & Smith, M. (2014). 双相情感障碍患者情绪和功能损害存在每周周期性的证据。《精神病学研究》, *218*(3), 290–294。Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). 生态瞬时评估的生态效度存疑：设计与分析的考量。《人类发展研究》, *14*(3), 253–270。（出版商：Routledge）Ryan, O., & Hamaker, E. L. (2022). 干预时机：网络分析与中心性的连续时间方法。《心理计量学》, *87*(1), 214–252。Samejima, F. (1969). 使用等级分数反应模式估计潜在能力。《心理计量学专刊增刊》, *34*, 100–100。Schueller, S. M., Aguilera, A., & Mohr, D. C. (2017). 针对抑郁症和焦虑的生态瞬时干预。《抑郁与焦虑》, *34*(6), 540–545。Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). 生态瞬时评估。《临床心理学年度评论》, *4*(1), 1–32。（出版商：Annual Reviews ） Shumway, R., & Stoffer, D. (2011). 时间序列及其应用。（期刊简称：附 R 示例，统计学 Springer 文本丛书；出版物标题：附 R 示例，统计学 Springer 文本丛书）Silverstein, A. (1988). 个体化与规律研究之间张力的亚里士多德式解决方案。《美国心理学家》, *43*, 425–430。（地点：美国；出版商：美国心理学会） Simon J. Julier, & Jeffrey K. Uhlmann. (1997). 卡尔曼滤波器在非线性系统中的新扩展。载于（第 3068 卷，第 182–193 页）。Slipetz, L., Falk, A., & Henry, T. (n.d.). 生态瞬时评估数据离散时间状态空间建模中的缺失数据：问题与建议。Sliwinski, M. J. (2008). 面向社会健康研究的测量突发设计。《社会与人格心理学指南》, *2*(1), 245–261。（出版商：John Wiley & Sons, Ltd）Srivastava, P., Michael, M. L., Manasse, S. M., & Juarascio, A. S. (2021). 身体不满的瞬时变化能否预测暴食发作？一项生态瞬时评估研究。《饮食与体重障碍——厌食症、贪食症与肥胖症研究》,*26*(1), 395–400。Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). 提高透明度

通过多宇宙分析。心理科学展望，*11*(5), 702-712. (PMID: 27694465) Stone, A. A., Schneider, S., & Harter, J. K. (2012). 美国每周各日情绪模式：论“忧郁星期一”、“谢天谢地，今天是星期五”以及周末效应的存在。《积极心理学杂志》，*7*(4), 306–314. Tuerlinckx, F., Maris, E., Ratcliff, R., & De Boeck, P. (2001). 对四种模拟扩散过程的方法进行比较。《行为研究方法、仪器与计算机》，*33*(4), 443–456. van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice：R 中的链式方程多重插补。《统计软件杂志》，*45*(3), 1 – 67.（章节：文章）van Rijn, P. W., & Molenaar, P. C. (2005). 单被试时间序列的逻辑斯蒂模型。载于《社会与行为科学中分类数据分析的新进展。(第 125–145 页)。美国新泽西州马瓦：Lawrence Erlbaum Associates出版社。Voutilainen, A., Pitk¨aaho, T., Kvist, T., & Vehvil¨ainen-Julkunen, K. (2016). 如何询问患者满意度？与对称李克特量表相比，视觉模拟量表较少受到混杂因素和天花板效应的影响。《高级护理杂志》，*72*(4), 946–957. Walani, S. R., & Cleland, C. M. (2015). 多重插补法：涉及二次数据分析的案例研究。《护理研究》，*22*(5), 13–19. Wright, A. C., Browne, J., Skiest, H., Bhiku, K., Baker, J. T., & Cather, C. (2021). 精神分裂症谱系障碍中常规临床评估与症状和功能状况的瞬时评估之间的关系：系统综述。《精神分裂症研究》，*232*，11–27。