软件过程仿真建模度量

Bohan Liu、He Zhang、Liming Dong、Zhiqi Wang、Shanshan Li

摘要—背景：软件过程仿真（SPS）已成为软件过程管理与改进的有效工具。然而，由于测量成本负担沉重且对领域知识的要求很高，其在工业界的应用程度低于研究界的预期。从过程实施的真实数据中提取恰当度量指标的难度，是面临的一大挑战。

目标：我们旨在为软件过程（仿真）建模中的过程度量指标提供循证支持。方法：我们扩展了此前的一系列综述，开展了一项系统性文献综述，以依据SPS度量指标本体的元模型，全面理解过程建模度量指标。

结果：我们确定了145项过程建模研究，这些研究共涉及2130个度量，并采用编码技术对其进行分类。我们针对两个层级的建模目的，提出了两张图表，用以展示度量指标之间高频使用的因果关系。我们比较了不同范式下的具体度量指标，主要区别在于：离散事件仿真（DES）和基于智能体的仿真（ABS）能够从开发活动和个人开发人员的角度提供更细致的仿真，而系统动力学（SD）则倾向于使用平均值作为替代。我们重新审视了SPS数据准备阶段遇到的数据问题，并确定了相应的应对策略。

结论：本研究的结果为过程建模人员提供了基于证据的参考，说明如何在SPS建模中识别和使用度量指标，并进一步推动过程建模背景下软件度量知识体系的发展。此外，本研究并不局限于过程仿真，还可扩展至一般的软件过程建模。将仿真度量指标作为标准和参考，可以进一步激励并指导软件开发人员在实践中改进过程数据的收集、治理和应用。

索引词—软件度量、软件过程模型、过程仿真、系统性文献综述

arXiv:2301.06390v1 [cs.SE] 16 Jan 2023

1 引言

SOFTWARE过程模型旨在深入了解软件过程，以便我们能够预测、修改或控制这些过程 [1]。软件过程模型可以是静态（描述性）模型，也可以是动态（仿真）模型，其行为会随时间变化。仿真需要更多信息和知识，但能够更详细地模拟现实世界。人们普遍认为并接受，SPS是支持软件过程管理和改进的有效工具。自Abdel-Hamid和Madnick [2] 于20世纪80年代将软件过程仿真（SPS）引入软件工程（SE）以来，该领域已发表大量研究，其中包括不少工业案例。Ahmed等人 [3] 开展了一项调查，以了解SPS仿真实践的实践现状；近半数受访者（8/17）来自工业界。此外，研究人员将SPS技术应用于能力成熟度模型集成（CMMI），以实现过程优化 [4, 5, 6],；SPS被认为是达到4级和 5 [7]的关键。除CMMI外，Mishra等人 [8]还构建了一个系统动力学模型，以了解印度软件行业的全球软件开发工业界。系统动力学技术也被应用于为新产品开发项目选择最佳的闸门时机策略 [Van 17][1]。

Zhang 等人 [9] 指出，SPS 在预测、过程调查、技术评估和风险管理等多种目的方面具有优势。为实现建模目的，SPS 模型可能涉及许多度量指标（有时甚至多达数百个）。识别特定过程模型所需的度量指标及其关系是一项具有挑战性的任务，尤其是对于新手建模人员而言；而收集这些度量指标的高质量数据更是耗费精力。

SPS领域专家组指出，在工业界构建SPS模型的前提是公司能够分析自身的信息需求 [10]。识别合适的度量指标并收集相应数据占据了大部分成本。分析信息以及识别度量指标，需要相当丰富的知识和技能。因此，软件过程社区鼓励开发知识库和模型库，并将通用过程度量指标集合作为关键组成部分，以释放可复用性 [10]。许多SPS研究指出了度量指标方面的问题，以下列出部分证据：

“对于某些相关数据，几乎不可能确定

• *B. Liu、H. Zhang、L. Dong、Z. Wang* 和 *S. Li* 均来自中国江苏省南京市南京大学软件学院新型软件技术国家重点实验室。电子邮箱：*bohanliu@nju.edu.cn*、*hezhang@nju.edu.cn*、*lm- dongmg@gmail.com*、*502022320013@smail.nju.edu.cn*、*lss@nju.edu.cn*

1. 我们采用一种独特的引文格式，以区分综述中的研究与其他参考文献。

现实项目中的必要信息……我们正在深入探讨如何将这些无法直接观测的人类属性纳入考量，并在模型的定量逻辑中考虑这些属性。”[Neu 03]。

“无论是否开发仿真模型，获取定量数据都是另一项难题。”[Park 07]

建模度量指标方面的挑战主要有两项：1）基于领域知识和现有数据，从真实过程识别关键度量指标；2）收集并挖掘测量所需的数据。了解SPS建模中使用了哪些度量指标，是研究这两项挑战的前提。据我们所知，尽管过去几十年间已有许多关于软件度量的论文和书籍 [11, 12, 13, 14, 15, 16, 17] 问世，但迄今尚无专门研究SPS建模度量指标的二次研究。这促使我们对SPS模型采用的度量指标进行循证梳理，以助力过程建模领域软件度量知识体系的发展。

因此，本研究及后续研究旨在系统识别示例过程模型中的建模度量指标，并从现有相关文献中提炼经验，从而减轻SPS建模的巨大负担和成本。尽管本研究以SPS建模中的度量指标为研究对象，但其也适用于一般的软件过程建模和软件过程测量。构建仿真模型通常包含以下步骤。在构建可执行仿真模型的过程中，静态模型通常是必不可少的中间产物。从评估角度来看，我们要求描述性模型在语义上正确；同时，由于仿真模型的输出是特定项目 [18, 19]的一组数值分布，因此我们需要评估其适当性和保真度。从建模角度来看，构建描述性模型就是根据迄今获得的不完整知识，对现实世界的片段进行抽象、收集和转换 [20]。仿真模型需要对所仿真的过程有详细的理解，并需要可靠的数据来进行初始构建。例如，需要指定一组变量，以表示时间的连续可微函数 [Zhan 06a，Zhan 09]。因此，与静态过程建模相比，SPS建模对度量指标的要求更高。SPS模型中使用的度量指标在大多数情况下也适用于静态模型。此外，还可以考察建模驱动下描述现实世界所需的度量指标。

为此，我们构建了一个用于建模的度量本体元模型，以指导本研究。遵循该元模型，我们开展了系统文献综述（SLR），对 SPS 模型中使用的度量指标进行汇总、分类和综合。截至 2021年，我们从 SPS 相关论文中筛选出 145 项报告了 SPS 模型的研究。我们从纳入的研究中提取了 2130个度量。尽管软件度量领域已提出软件度量分类方案 [21, 22], , 但由于 SPS 研究与软件度量研究的重点不同，这些方案并不适用于 SPS建模。

在该元模型的指导下，我们围绕四个研究问题，考察了度量指标及其直接相关的要素。我们基于提取的度量指标，并参考该研究建议的高层类别（实体和属性），构建了一个新的分类框架 [22] （研究问题1：度量指标）。我们讨论了SPS模型中度量之间的因果关系，并分别研究了 SPS模型的建模目的和范式在度量指标选择方面的考量（研究问题2：度量之间的因果关系）。我们提出了因果关系图，用以说明认知层级模型与战术与战略层级模型在建模方面的异同，并讨论了不同范式所用度量指标的差异（研究问题3：度量指标的筛选）。我们还映射了测量中的数据问题与解决策略之间的关系（研究问题4：度量指标的数据）。

本研究对研究界和工业界的从业者都有贡献。

- 本研究应对了建模中的第一个挑战（识别关键度量指标）；同时，为研究人员开展研究以解决第二个挑战（获取测量所需的数据）奠定了必要基础。
- 本研究可作为知识库，使从业者和研究人员能够全面了解SPS建模所涉及的度量指标及其关系。
- 从建模目的、范式和数据问题等方面对度量指标进行分类的框架及相关考量，可为从业者复用现有知识提供参考；同时，也为研究人员指出了需要重点关注的度量指标及因果关系。
- 本研究基于证据讨论了第二个挑战，并识别出一系列数据问题、应对策略，以及难以获取的度量指标所对应的可用数据来源。

请注意，我们采用一种不同的引用格式（作者姓氏和年份，例如 [Pfah 01]），以便将综述中的研究与其他参考文献区分开来。纳入的研究的完整参考文献列表见附录。

2 相关工作

本节介绍软件工程界数十年来一直研究的过程仿真和软件度量指标。

2.1 软件过程仿真

Kellner 等人 [1] 概述了软件过程仿真领域，旨在回答三个基本问题，即为什么、是什么以及如何做。他们指出了开展软件过程仿真研究的原因，界定了SPS模型的范围，并讨论了目的、范围与度量指标之间的关系。他们还提出了一个框架，用于支持对仿真方法、技术和所需度量指标的判定。

Zhang 等人 [23, 24, 25] 围绕 SPS 建模开展了一系列系统性文献综述。他们确定了 SPS 研究的十种目的，并将其划分为三个级别，以及模型范围的两个维度 [23]。此外，还总结了若干建模范式和

仿真工具 [23, 25]。他们根据研究结果指出了过程建模的五种趋势 [24]。另一项研究 [26] 还报告了 SPS 研究对实践的影响，并提出了一份影响路线图，追溯成功的 SPS 工业应用案例的起源。

过去十年间，SPS的影响力逐渐扩大，也日趋成熟。[27], 研究了SPS在软件工程教育中的应用，证实教育是SPS的重要应用领域，持续受到研究关注，并表明与其他形式的SPS相比，SPS游戏似乎对教育工作者更具吸引力。随着软件过程变得日益复杂，研究者制定了构建混合SPS模型的集成策略和建议 [28]。验证与确认（V&V）在确保SPS模型质量方面发挥着关键作用， [29]提出了用于验证与确认的质量要素与可能采用的验证与确认方法之间的映射。

然而，关于SPS研究的影响和实用性，仍存在争议。Franc.a等人 [30]对108项研究进行了准系统综述，以调查SPS研究在软件工程中的可靠性。结果，他们发现了一些问题，并指出SPS研究缺少复现所需的必要信息。Ali等人 [31]汇总了关于SPS实用性的各种观点，但发现根据二次研究，无法对这些相互矛盾的主张得出结论。他们开展了一项系统文献综述，并评估了87项SPS研究。结果表明，目前仍缺乏确凿证据，因为只有少数研究报告了开发SPS模型的成本。Pfahl [32] 也指出，仍缺乏证据证明SPS已成为软件项目经理认可并定期使用的工具，而其高昂成本是主要原因。另一方面，SPS领域专家组共同提出了不同的观点：与许多其他软件工程技术相比，SPS对实践的影响不容忽视；尽管高成本仍是其广泛应用的主要障碍，但应考虑放弃仿真所带来的后果，而在其他工程学科中，仿真被视为成本节约工具，而非成本驱动因素 [10]。

2.2 软件（过程）度量

软件度量在定量软件工程中发挥着至关重要的作用，多年来一直从不同角度得到研究。我们讨论了以往关于软件度量的二次研究，并在表1中进行了概述。

G´omez 等人 [11] 进行了系统性文献综述，旨在回答如何、何时以及测量什么等问题。他们采用了《软件度量本体》中定义的概念分类。该本体由 [33] 提出，旨在促进不同软件度量方案和标准的协调统一，并提供一套用于软件度量的通用概念。

Bellini 等人 [12] 开展了一项系统文献综述，旨在探究五个关键的概念与方法问题，即如何将测量理论应用于软件、如何构建软件度量、如何开发度量指标、如何收集核心度量指标，以及如何分析度量指标。他们采用了 Fenton的分类框架 [22]。最终，他们提出了收集和分析测量结果的方法，并指出人们正日益关注多维度量。

Kitchenham 等人 [13] 初步绘制了软件度量研究的映射图，重点是识别 2000 年至 2005年期间有关软件度量的影响力研究。他们得出结论：实证研究对软件度量研究群体至关重要。尽管许多研究人员从不同维度对软件度量进行了研究，但仍需改进所采用的实证方法。

Abilio等人 [14] 开展了一项系统文献综述，从 11 篇主要论文中识别与软件可维护性相关、并针对面向特征和面向方面技术提出的度量指标。这些度量指标根据其所测量的软件属性进行分类，涵盖架构、并行开发、调试和质量属性。

Kupiainen等人 [15] 针对工业环境中的精益和敏捷软件开发中度量指标的使用开展了一项系统文献综述。作者依据Fenton分类框架 [22]对度量指标进行分类。他们识别出度量指标在敏捷开发中的影响程度和流行度，并展示了度量指标与敏捷原则之间的映射关系。

Nu˜nez-Varela 等人 [16] 开展了一项映射研究，以调查代码度量研究的发展趋势。他们将代码度量分为四类，包括面向对象编程、面向方面编程、面向特征编程和过程式编程。

Meidan等人 [17] 开展了一项映射研究，旨在创建软件开发过程测量研究的分类方案，分类依据包括来源类型、出版年份、研究类型、贡献类型、提案类型、验证类型、实体/抽象和研究背景。他们识别出 13 个过程属性、3 个开发者属性、4 个项目属性和 2 个组织属性。

本研究的范围各不相同，我们更关注过程度量指标，这有助于SPS研究人员全面了解如何在软件过程（仿真）建模中使用度量指标。

3 本体

为了系统地研究SPS建模的度量指标，我们提出了一个本体，用以阐明度量指标与仿真建模之间的关系，以及相关概念。我们的研究围绕这一本体展开。

Olsina和Mart´ın [34] 基于不同的软件相关ISO标准和研究文章，提出了一个关于软件度量指标和指标的本体。为了适用于SPS，我们参考其本体中度量指标部分，提出了一个调整后的本体元模型，并引入了SPS建模这一概念，如图1所示。元模型中所述概念及其关系的描述分别见表2和表3。

所有模型都是对现实世界的简化表示（抽象），而概念模型则是仿真模型的抽象。所有仿真建模都涉及概念建模 [35]。概念建模

表1： 本研究与以往软件度量研究之间比较的概述

研究	Year	度量指标的分类	研究重点
G´omez 等人 [11]	2006	软件度量本体 [33]	测量什么、何时测量以及如何测量
Bellini 等人 [12]	2008	Fenton 分类法 [22]	测量理论；收集和分析核心
Kitchenham 等人 [13]	2010	面向对象度量、Web 度量及其他代码度量	度量的方法；度量概念，以及如何识别度量指标 识别有影响力的软件度量研究中的趋势
Abilio等人 [14]	2012	论文中测量的软件属性	软件可维护性度量
Kupiainen等人 [15]	2015	Fenton 分类法 [22]	在敏捷软件中使用度量指标
Nu˜nez-Varela等人 [16]	2017	面向对象 编程、面向方面 编程、面向特征编程以及 过程式编程	源代码度量的发展趋势
Meidan等人 [17]	2018	过程、开发人员、项目和组织	了解软件开发过程的测量
这项工作	2022	基于 Fenton等人的详细框架 分类法 [22]	针对软件过程建模中的度量指标，包括分类—— 化、度量之间的因果关系，以及建模中的度量指标筛选 和测量的数据问题。

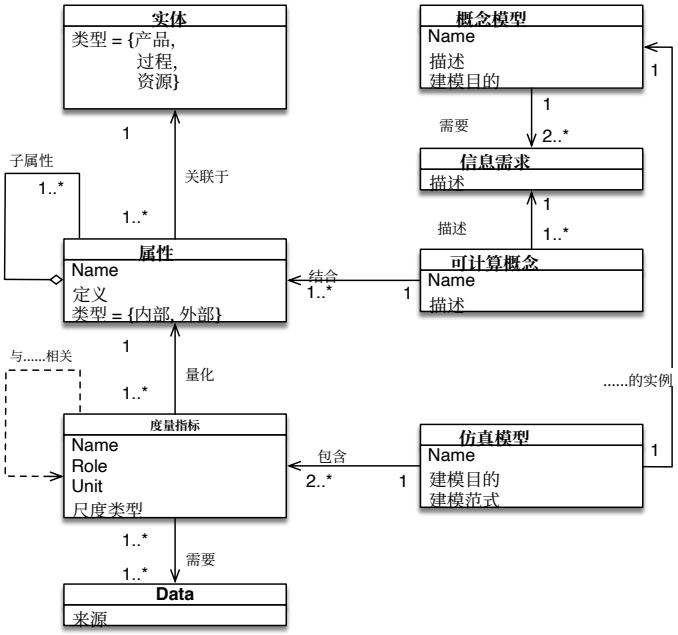

图 1：软件过程仿真建模中软件度量本体的元模型

是指指定一个模型，以表示问题领域中纳入仿真模型的部分。根据 Kitchenham 等人提出的模型开发过程 [19]，，在指定模型之前，需要进行领域分析并定义模型需求。领域分析旨在确认现实世界中的信息需求。并非所有信息需求都应纳入模型需求，因为这些需求应当有效、可信、可行且有用 [36]，，换言之，它们是可计算概念。仿真模型通过特定的计算机代码表示概念模型，这意味着仿真建模旨在量化概念模型的输入和输出。从软件度量的角度来看，它旨在测量实体的属性，并量化属性之间的因果关系。

4 研究方法

本节介绍了关于SPS建模中度量指标的系统文献综述过程，该过程遵循系统文献综述指南 [38]。四位研究人员及其导师参与了这项研究。

严格来说，这项研究并非系统文献综述研究。我们采用系统文献综述研究方法来研究度量指标，而这些指标通常不是我们纳入的原始研究的研究重点。

4.1 研究问题

本研究聚焦于元模型（图 1）中与度量指标相关的度量指标及其要素。研究问题1从不同角度研究度量指标及其相应属性。研究问题2研究SPS模型中呈现的度量指标之间的因果关系。研究问题3研究SPS模型的目的和范式，这些因素会影响度量指标的筛选。研究问题4研究测量数据的问题及解决策略。作为概念模型的实例化可执行模型，仿真模型能够涵盖概念模型的相关信息。因此，研究问题并未专门针对元模型中涉及的概念模型、信息需求和可计算概念。为实现研究目标，本研究依照元模型（如下所示）提出四个研究问题，以指导本研究。

研究问题1： ***SPS***研究中使用和研究了哪些度量指标？研究问题1旨在深入探讨需要度量什么，并构建一个分类框架，对SPS模型中使用和研究的软件度量进行分类。如元模型所示，度量指标用于量化与实体相关联的属性。我们还将对度量指标进行多级分类，根据其所度量的属性及与该属性对应的实体对其分组。此外，还将列出每个度量指标的单位和尺度类型，以更具体地说明如何进行度量。

研究问题2： 不同***SPS***研究中所用度量指标之间存在怎样的因果关系？研究问题2有助于识别现有SPS模型所用软件度量之间的常见因果关系。仿真模型由度量指标之间的因果关系构成。度量指标之间的因果关系是理解SPS模型结构的基础。

研究问题3： 针对不同的建模目的，选择度量指标时应考虑哪些因素？

几乎不可能在一个模型中仿真所有因素和细节。建模人员会选择最相关的度量指标，并忽略其他指标，因为他们关注的建模目的各不相同。此外，针对不同的建模粒度，建模人员会采用不同的建模范式，这会影响度量指标的测量。研究问题3的目标是

表2：软件度量本体：概念术语表

概念	定义	属性：描述
实体	要通过测量 其属性 [37] 来描述的对象。	名称： 实体的名称。 描述：对实体含义的明确描述。
属性	可测量的物理或抽象属性 的实体。	名称： 属性的名称。 定义： 对属性含义的明确描述 *[37]*。 类型：属性可以是内部的，也可以是外部的。
度量指标	度量指标是分配给某个对象的数量或符号 通过此映射关联一个实体，以便 描述某个属性。 [22] 变量和参数是 SPS模型中度量指标的同义词。	名称： 度量指标的名称。 单位： 按照惯例定义和采用的特定量，用于衡量其他同类量 同类数据会进行比较，以表达其相对于该量的大小。 尺度类型： 尺度的类型取决于尺度值之间关系的性质。 通常定义以下尺度类型：名义型、序数（受限或不受限）、区间型、 比率 ，和绝对型。 型
Data	在现实中需要收集的数据 用于测量的范围。	来源： 可以获取数据的来源。
概念模型	与软件无关的描述，内容涉及 计算机仿真模型（将要、 正在或已经开发），描述 目标、输入、输出、内容、假 设和模型 [36] 的简化。	名称：概念模型的名称。 建模目的：与现实世界相对应的特定目的，即我们开发模型的原因。
信息需求	特定建模所需的洞见 用途 [37]	描述：对信息需求的明确文字说明。
可计算概念	实体属性与 信息需求之间的抽象关系 [37]。	名称： 可计算概念的名称。 描述：对可计算概念含义的明确描述。
仿真模型	仿真模型是一个计算机化的 模型，用于表示某个动态 系统或现象 [1]。	名称： 仿真模型的名称。 建模目的： 与现实世界相对应的特定目的，即我们开发模型的原因。 建模范式： 用于仿真建模的一组独特概念或思维模式，例如： 系统动力学。

表 3：软件度量本体：关系描述

关系	描述
子属性	一个属性可以由零个或多个子 属性 ，构成，而这些子属性本身也是属性。
关联对象	一个或多个可测量属性与之关联 一个或多个实体。
量化	一个或多个度量指标可以量化一个属性。
需要	测量度量指标需要现实世界数据。
结合	可计算概念组合（关联）一个或多个 可测量属性。
由以下部分组成	仿真模型由若干度量指标组成。
以下各项的实例	仿真模型是一个概念模型的实例， 模型。
与.....相关	一个度量指标与另一个度量指标相关。

调查度量指标在建模目的和范式方面的考量；正如元模型所示，这两者是不同仿真模型的两个主要属性。

研究问题4： 获取度量指标数据时存在哪些问题和策略？数据缺乏仍然是 SPS 建模中的一个问题，因为 Raffo 和 Kellner [39]分析了不同情形及解决方案。本研究问题4旨在基于证据重新审视数据问题的现状，同时调查现有的解决策略以及这些策略可衡量的具体属性。

4.2 筛选标准

相关研究的纳入和排除标准如表4所示。纳入的论文应以英语发表，检索时间范围截至2021年。我们的研究纳入了将仿真建模范式应用于软件过程研究、软件教育和软件实践的研究。由于度量指标是本综述的重点，入选研究应明确说明其模型中使用的度量指标。常规论文可以为我们提供

为了获得更全面的信息并帮助我们收集重要证据，我们不会选择页数不超过 5 页的研究。研究应以期刊论文或会议论文的形式发表，而非 C5 和 C6 中所列的其他形式。我们从原始研究中提取数据，因为这些研究能够提供 SPS研究所使用的一手度量指标，而不是从相关的二次研究中提取。此外，原始研究中的证据有助于我们总结不同 SPS研究中各项度量指标之间的关系。另外，我们没有发现任何调查 SPS所用度量指标的二次研究。具体而言，尽管 Pfahl [40] 提出了一组度量指标示例，用于旨在支持分析汽车行业关键软件过程改进有效性的系统动力学模型（满足 C1、C2、C3），但该论文因篇幅较短而被排除（符合 C4）。另一个例子是，Zhang 等人 [41] 将四种典型仿真范式映射到适当的能力成熟度模型集成成熟度级别，以便采用这些范式；然而，该论文未提出具体模型（不满足 C2、C3）。

表4：筛选标准

纳入标准
C1. 发表于2021年之前，且以英语撰写。
C2. 关于在软件 过程研究、教育和实践中采用仿真建模范式的原始研究。
C3. 声明使用了仿真模型中的度量指标的原始研究。
排除标准
C4. 短篇论文（不超过 5 页）。
C5. 社论、摘要、主题演讲、海报和书籍等形式。
C6. 观点文章、评论、更正、说明、单独的幻灯片或立场 论文。
C7. 总结现有研究结果的二次研究 工作，例如路线图、综述、调查等。

4.3 检索与筛选流程

图2展示了本研究的检索过程，共分为五个阶段。第一阶段和第二阶段已在我们的初步综述中完成 [28]。在第一阶段，两名研究生分别进行了人工检索和自动检索。Gao等人的研究 [28]列出了人工检索所涉及的会议和期刊，包括五个会议和六种期刊。检索字符串见图2。随后，我们将其编码为相应形式，以匹配不同数字图书馆的检索语法。我们检索了四个数字图书馆（IEEE Xplore、ACM数字图书馆、ScienceDirect、SpringerLink），以尽可能多地获取 SPS 研究。第二阶段采用前向滚雪球法作为补充 [28]。最终，通过扫描标题、关键字和摘要，共确定了 331 项候选研究。

本研究与 Gao等人 [28] 的研究在筛选标准上的唯一差异是 C3：该标准从所有相关的 SPS研究中筛选出报告建模度量的研究（而非前一项研究 [28]中的混合仿真建模）。因此，在第三阶段，我们通过反复阅读引言、结论，甚至全文，依据筛选标准对331项候选研究进行核查，直至达成最终共识。候选论文分配给四名研究生，每篇论文至少由两名学生负责，以便并行综述。所有不一致和分歧都在每周与导师举行的会议上经过充分讨论并予以解决。在本研究中，如果基于同一模型发表了不同版本的原始研究，则只选取其最新版本进行综述。

在第四阶段，我们使用相同的检索字符串重复自动检索，并将检索范围扩展至2021年。我们确定了19篇发表于2016年至2021年且符合筛选标准的论文。

在第五阶段，我们使用 Google Scholar 进行了前向滚雪球法。前向滚雪球法的种子集包括“Abdel91” [2]，、“Kellner99” [1]，和 “Zhang10” [25]，这些文献均被综述早期阶段中大多数相关研究引用。去重后，我们确定了 5 项研究。

4.4 数据提取

为回答研究问题，我们定义了一套数据提取方案，用于收集所综述研究中的重要信息。每个研究问题均通过

至少一个提取项。如表5所示，数据提取方案包括引文信息（例如标题、年份和建模范式），以及与研究问题相关的具体信息。提取软件度量（名称、单位和原始描述），以回答研究问题1、2、3、4。识别度量指标（两项）之间的因果关系和建模目的，分别用于回答研究问题2和研究问题3。识别数据来源、问题及解决方案，以回答研究问题4。

我们首先进行了试点提取，将随机选出的35篇论文分配给所有研究人员。我们注意到，有些论文明确介绍了用于SPS建模的度量指标，或展示了SPS模型。在这些情况下，我们可以轻松识别软件度量及其因果关系。有些论文没有明确列出软件

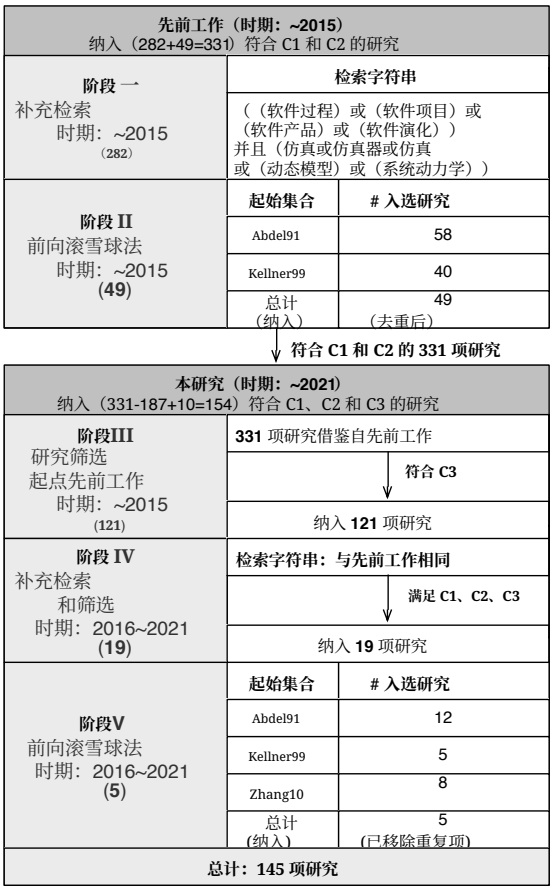

图2：全面文献检索流程

表5：数据提取方案

Item	描述	研究问题
标题	研究标题。	Info
Year	研究发表年份。	Info
范式	系统动力学；离散事件仿真——；基于智能体的仿真；混合仿真等。	Info
软件度量	SPS研究中涉及的度量指标，包括其名称、单位和原始描述。	研究问题1-4
因果的关系之间度量指标	输出到或影响当前指标的前驱度量指标 影响当前指标，以及后续的作为输入或受当前指标影响的后续指标。	RQ2
建模目的	构建 SPS 模型的原因。	RQ3
数据问题	因以下原因而选择度量指标所产生的问题 数据可用性。	RQ4
数据问题的解决方案是——sues	解决数据问题的策略。	RQ4
数据来源	与解决方案相关的数据来源，用于测量属性，一项研究可能会使用多个数据来源。	RQ4

度量指标；通过反复阅读全文，或许可以根据上下文识别出这些度量指标及其因果关系。最终，我们从已识别的145篇论文中提取了2130个度量，并识别出183种因果关系类型。这些度量也称为变量，构成了SPS模型。

4.5 数据综合与分类

我们采用主题综合法 [42] ，以系统的方式构建研究发现。由于度量指标

表6：编码示例

度量指标	Code-I1	Code-I2	Code-I3	Code-I4
E1: 残留缺陷密度（即实际已报告但1094天后仍未纠正的缺陷）	缺陷，密度，残留	缺陷，密度，剩余	缺陷，密度	缺陷
E2: 已完成的任务数量	任务，数量，已完成	任务，数量，已完成	任务，规模	Task
E3: 从完成此任务到下次发布的延迟为交付给用户	延迟、发布	延迟、发布	延迟、发布	Time

描述在不同过程建模人员之间存在显著差异，为回答研究问题1，我们使用编码技术 [43]对提取出的软件度量指标列表进行综合。这是一个迭代过程，旨在从综述研究中多样化的描述中形成一套一致的代码。表6展示了代码演变过程中的三个示例。在第一轮迭代中，两位研究人员根据描述，分别对所有度量指标进行细致编码；随后，我们讨论两组代码之间的差异，最终达成一致。在第二轮迭代中，我们找出第一轮形成的Code-I1中的相似代码，并将其统一为相同代码。例如，我们将“问题”“错误”“故障”“缺陷”“程序错误”编码为“缺陷”；将“修复”“已修复”“纠正”“更正”编码为“修复”。对于示例1（E1），Code-I1中的“残留”被Code-I2中的“剩余”所替代。有些代码没有可供替换的同义词，因此Code-I2与Code-I1相同，例如E2。经过讨论，我们在第三轮和第四轮迭代中逐步提高代码的抽象程度。从I2到I3，有些代码并未发生变化，也没有用于替换相似的度量指标，例如E3。我们基于Code-I3构建了Code-I4，并部分参考了Fenton等人提出的软件度量分类体系 [22] ，该体系已在学界得到广泛认可。因此，根据Code-I4，所有度量指标均可分为29个不同类别；按照该研究 [22]的建议，这些类别可归为产品内部/外部、过程内部/外部和资源内部/外部。我们依据Code-I3将这些类别进一步划分为子类别。

在对度量指标进行基于编码的分类过程中，可以轻松识别并分类衡量时间、工作量和成本等的度量指标。但我们也遇到了一些棘手的问题，这些问题在 Fenton等人的工作 [22]中并未涉及。模型中存在各种因素和乘数，会影响其他多种变量。根据受其影响的变量，我们将这些因素分为两大类，即过程因子和人力因子。例如，现有的SPS模型对各种政策进行了建模；我们将结对编程策略归为过程因子，它是一个决定是否采用结对编程的布尔变量；并将人员配置政策 归为人力因子，它决定分配的人月数量。在Fenton等人的框架 [22],中，缺陷 是唯一与缺陷相关的类别，他们建议将缺陷作为过程属性。我们认为，尽管缺陷是在过程中产生和修复的，但在某一时点存在的缺陷是客观存在的，是产品的一部分。因此，我们将缺陷活动 与缺陷 区分开来，并将缺陷归入产品类别，这样更为合理。我们将缺陷修复率等度量指标归入缺陷活动 （属于过程），并将# 已修复等度量指标 缺陷 归入缺陷。同样，我们一方面根据编码的结果（即度量指标的抽象含义）进行分类，另一方面也依据度量指标的上下文进行分类。我们参考了现有的分类方法，但并不受其限制。

在整个研究过程中，我们进行了内容（频率）分析，以分析不同变量组中的名义数据。为了确定度量指标之间最常见的因果关系，在回答研究问题2时，我们统计了因果关系出现的百分比。

对于研究问题3，我们统计了用于两个层级建模目的以及不同范式的各类度量指标的频率，以找到导致差异的证据，从而开展深入分析。

5 结果

本节从三个方面展示纳入的研究的分布，即发表年份、研究中使用的范式以及建模目的。

5.1 年份

经过五个阶段后，综述确定了145篇论文（如图2所示）。选用于综述的论文发表于1997年至2021年。纳入的研究的完整参考文献列表见附录。

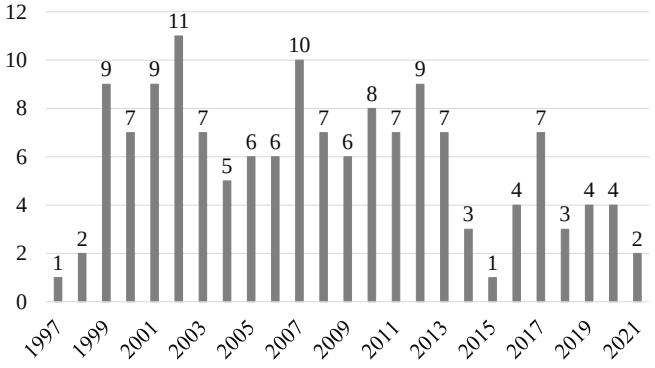

图3：各年份的研究分布

5.2 范式

我们对 SPS 研究中采用的建模范式进行了分类，如图 4 所示。系统动力学（SD，43%）是 SPS 中最常用的建模范式。其余包括离散事件仿真（DES，19%）、混合仿真（混合，12%）和基于智能体的仿真（ABS，8%）。混合模型可以采用两种或更多种建模范式

。其他研究采用了多种建模范式，以不同于上述范式的抽象层次来仿真软件过程，例如定性仿真（QSIM）、参数估算（PE）等。

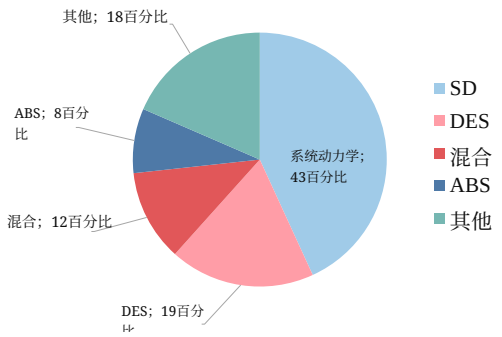

图4：各建模范式的研究分布

5.3 建模目的

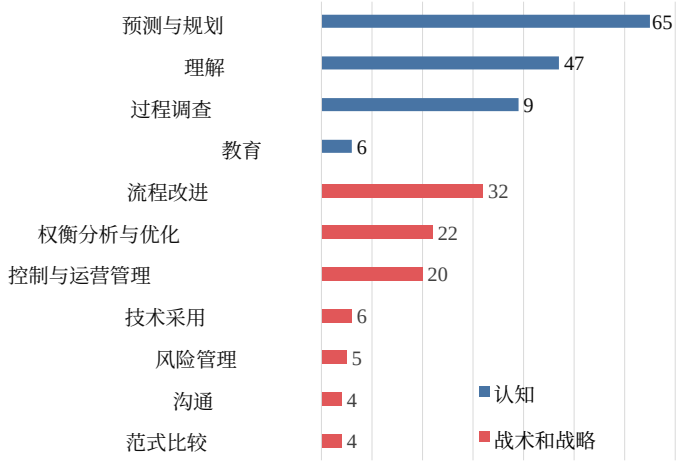

图5：各建模目的的研究分布（同一项研究可能有多个建模目的）

Kellner 等人 [1] 指出了 SPS 模型的用途，并将其归纳为六种具体的建模目的，以便对软件过程进行仿真。Zhang 等人 [25] 基于对已发表 SPS 研究的系统综述，将其扩展为十种目的。在他们的研究中，这些目的被归为三个级别，即认知、战术和战略。认知层级包括理解、沟通、过程调查和教育。其余目的则属于战术和战略层级。

我们将这些研究的建模目的归纳为十一个类别（我们识别出一种新的建模目的）。与其他十种目的不同，新的目的——范式比较——与建模范式有关，而不是为了研究软件过程。其目的是比较不同范式的优势与劣势。在目标为比较范式的研究中，范式比较是构建模型的唯一目的。例如，为了比较系统动力学与定性建模图，研究人员分别构建了典型软件演化过程的定性模型和定量模型，而且没有多余地讨论模型本身的价值 [Zhan 09]。为范式比较而构建的模型较为简单，但足够完整，其依据是模拟行为对过程的解释程度。从这一角度来看，这些模型处于认知层级。

如图 5 所示，我们将 42 项研究归入认知层级，其余 103 项归入战术和战略层级。一项研究可能有多个建模目的，因此可能同时具有认知层级和战术、战略层级的建模目的。对于此类研究，我们将其建模粒度视为战术和战略层级。理解和过程调查是认知层级最常见的建模目的。预测与规划以及流程改进是战术和战略层级最常见的建模目的。

6 研究发现

Kaner 等人 [44] 收集了测量的定义，他们推荐采用 Fenton 和 Pfleeger 提出的简明定义，如下所示 [22]。

形式上，我们将测量定义为从经验世界到形式化关系世界的映射。因此，度量指标是通过这种映射赋予某个实体的数字或符号，用于表征某项属性。

在 SPS 建模中，度量指标有时也称为变量。我们认为，度量指标是更恰当的名称，因为它意味着从真实的软件过程到 SPS 模型的映射。然而，在 SPS 研究中，属性与度量指标之间的界限并不严格。例如，代码大小 和缺陷数量 是两个变量的常见名称，但根据软件度量领域的定义，前者是属性，后者是度量指标。代码大小 的度量指标应为代码行数。这种歧义会给分类带来麻烦。因此，我们将所有变量都视为度量指标，并尽可能保留其原始表述。

尽管软件度量方面已有一定的知识积累， [22], 但软件工程界尚未报告过针对过程建模度量指标的系统性、全面性研究。由于动态特性以及对可执行性的额外要求，SPS模型中的度量指标筛选比静态模型更具挑战性。因此，本研究仅聚焦并报告SPS建模中使用和研究的度量指标。

6.1 度量指标及其分类（研究问题1）

我们从综述的145项研究中，提取出SPS模型中使用（或研究）的2130个度量。数据集中存在许多相同或相似的度量，因此我们对其进行了分类分析。基于元模型，我们的分类框架包含四个由抽象到具体的类别层级，分别是实体、属性、子属性和度量指标。

6.1.1 实体类别

我们参照Fenton等人的分类及定义 [22]。我们将软件度量分为三类：产品、过程和资源 （与软件度量的基于目标的框架相一致）。其中，过程 是随时间演进、并且在开发过程中必须按顺序完成的活动；产品是由过程活动生成的，包括制品、交付物和文档；资源 则是执行活动所需的实体。

在每个类别中，我们进一步将度量指标区分为两种类型，即内部度量指标和外部度量指标。内部度量指标只能通过产品、过程或资源本身的内在属性来测量，而不考虑它们的可观察行为。相反，外部度量指标的测量完全考虑其可能对产品、过程或资源产生的影响。

6.1.2 属性类别

软件度量/测量领域已有大量研究，然而，度量指标在SPS研究中的作用与其在这些研究中的作用不同，这很容易使建模人员感到困惑。例如，软件测量相关研究中充斥着关于耦合和内聚的研究，而这些内容在现有SPS模型中却很少使用。这表明，基于软件测量研究和知识建立的现有分类框架并不适用于SPS建模。此外，最受认可的框架之一——Fenton等人的分类框架，并未完全涵盖建模中使用的度量指标。在研究 [22],中，每种实体仅列出了属性示例，这远不足以指导建模。

本研究按照第 4.5 节所述的主题综合法，构建了一个用于 SPS 模型的度量指标新分类框架。由于单页篇幅有限，该框架分为四幅图，即图 6–9。在这些图中，第一列显示所测属性属于内部实体还是外部实体。括号中的数字表示测量这些属性的度量指标数量。第二列和第三列分别显示属性及其子类别。最后一列列出典型度量指标及其单位。每个度量指标由两部分组成：上方的矩形显示度量指标的名称，下方的矩形显示其单位。例如，图 6 中的第一个度量指标是代码大小，其单位可以是代码行数、DSI 或功能点（功能点）。度量指标矩形的高度对应于其所属子类别的范围，例如，代码大小 测量制品规模，具体而言是代码规模。需要注意的是，从语义角度看，代码大小应是属性名称；然而，这些是 SPS 模型中的最小单位。我们按照编码技术为这些度量指标命名，并遵循综述研究中的原始表述。为实现建模目标，不同研究采用了多种方式（以单位表示）来测量特定度量指标。度量指标的单位可能不同，这在很大程度上取决于建模对象和测量方法。若将所有相似的度量指标（例如，代码行数和 DSI 都是代码规模的度量指标）分别列出，会使结果和讨论过于琐碎。因此，我们省略了一些非必要的结果。

单位也体现了度量指标的尺度类型，尤其是对于数值数据。如表 2 所示，根据 ISO/IEC 15939，尺度类型包括名义型、序数型（受限或不受限）、区间型、比率型和绝对型。绝对型和比率型属于数值数据，可用单位表示。例如，# 缺陷 属于绝对型，而缺陷 百分比属于比率型。此外，对于布尔型、区间型和序数型，由于这些度量指标没有维度（即没有特定单位），因此会显示类型名称而非单位。对于具有明确序数集合或范围的度量指标，也会显示该序数集合或范围。例如，开发者的专业水平通常有 0、1、2 三个级别，而知识级别 的粒度更细，范围为 0 到 100。

图中所示的分类框架按产品、过程和资源 类别进一步阐述如下。

6.1.3 产品度量

如图 6 所示，内部度量进一步分为三个子类别，即制品规模、制品属性 和缺陷。

制品规模度量过程中要产出的工作产品的数量。通常，规模用于衡量项目的大小，并进一步计算间接属性。不同研究采用了针对各种类型制品的度量指标，收集到的度量指标可分为五个子类别，如图 6 所示。代码大小用于度量程序规模，是所有规模度量指标中最常用的一项，并与开发和维护的工作量以及产生的故障相关 [22]。此外，需求文档大小、设计文档大小和测试用例数量（#）也广泛用于 SPS模型。具体而言，测试用例通过手动或自动化方法生成；后者考虑了支持基于模型的测试的工具，并基于以更正式的符号或结构编写的软件规格说明 [Aran 08]。在 [Wern 02, Zhan 12a], 等研究中，制品规模被抽象为若干大小任意的 “ 单位 ”，以表示对系统实际规模的某种适当度量。用户故事数量（#）和功能数量（#）用于不同的项目情境。

制品属性用于以更精细的级别度量产品；例如，可以进一步通过代码中的判定语句数量、局部变量和全局变量的数量等度量代码，也可以通过每页的字数等度量文档。具体而言，骨架中识别出的控制流数量用于估算在 DES模型 [Aran 08]中执行需求分析活动所需的工作量。耦合度和代码复杂度也属于此类别。前者仅用于一项研究，通过一个组件受另一个组件影响的概率来测量 [Padb 11]。后者可以通过不同方式测量。Smith等人 [Smit 06] 使用麦凯布复杂度来测量代码复杂度。Raffo等人 [Raff 00] 建议将代码中的判定语句数量作为代码复杂度的度量指标，局部变量和全局变量的数量以及控制流嵌套级别则是替代度量指标。建议使用 *Flesch-Kincaid*年级水平分数和*Gunning Fog*指数作为文档复杂度的度量指标。

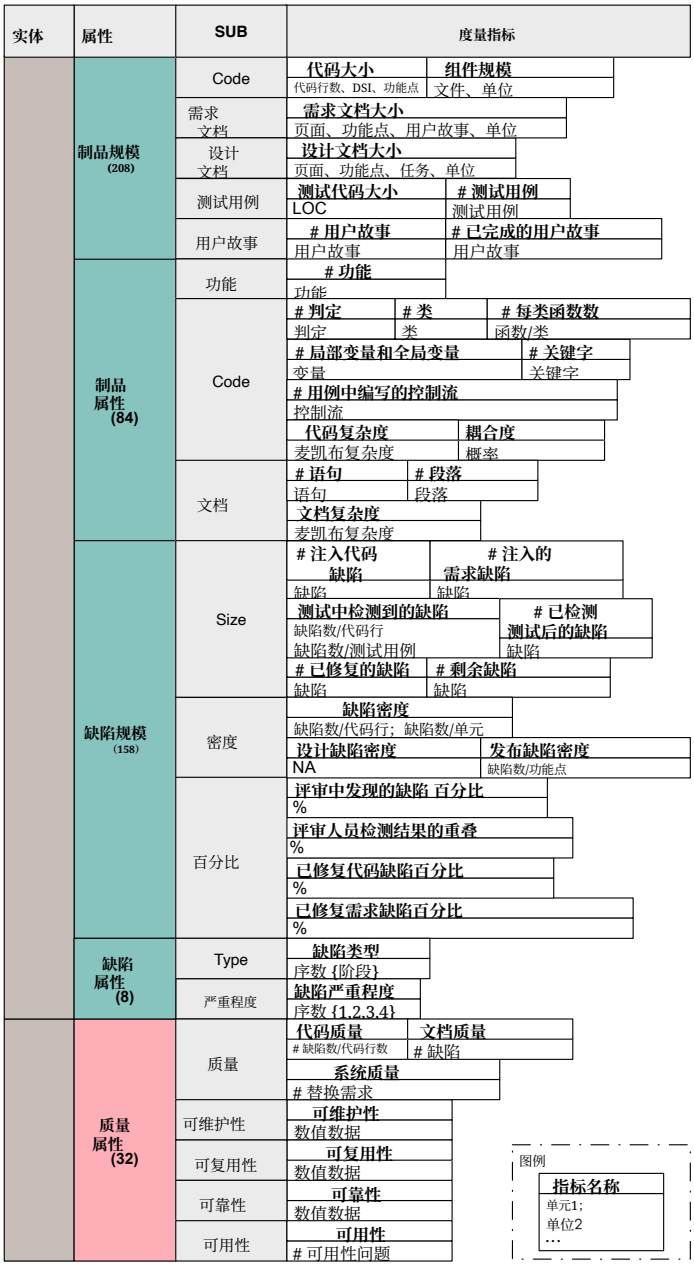

图 6：产品度量的分类

缺陷规模用于衡量在开发过程中注入、已检测以及可能仍然存在的缺陷数量。根据缺陷生命周期，缺陷大致包括注入的、已检测、已修复和剩余的缺陷。从测量角度来看，这些度量指标分为规模、密度、百分比和属性几个子类别。缺陷数量是最常用的类型。缺陷密度和已修复缺陷的百分比从两个不同角度评估产品质量：前者从产品本身进行评估，后者则从缺陷修复进度进行评估。

缺陷属性 用于以更细致的级别衡量缺陷。在多项研究中 [Rus03、Zhan08a、Cang 08、Lune 21] 引入了类型 和严重程度 等属性，使模型更贴近现实。类型 可以是顺序数据，用于指示缺陷注入的阶段 [Rus 02]。严重程度 可以是具有三到四个级别的顺序数据，例如，Zhang 等人 [Zhan 08a]将缺陷严重程度建模为四个级别（即“容易”“中等”“困难”和“非常困难”）。

在SPS中，产品的外部度量并不常用。它们衡量产品的质量属性，例如可复用性、可维护性、质量等。尽管质量是该领域常用的度量指标，但不同模型衡量代码和文档质量 的方法各不相同。Pfahl和Lebsanft采用系统动力学模型分析软件需求波动性的影响，并通过替换需求的数量来衡量系统质量 [Pfah 00b]。从业务价值的角度出发，Madachy [Mada 05] 通过缺陷数量来衡量质量 。只有一项研究使用了可维护性 度量指标，并用公式对其进行衡量 $Maintainability = 13.12 * Complexity + 0.17 * Effort + 3.87 * Size$ [Arau 12]。可复用性 和可靠性 通常基于COCOMO II，并作为成本驱动因素，其值为名义型 1 [Uzza 13b, Silv 13]。

6.1.4 过程度量指标

聚焦于过程建模，图7、8表明，过程相关度量指标的集合很大，还可以进一步分为若干类别，例如时间、任务、工作量等。

作为动态过程模型与静态过程模型之间的区别，时间度量指标共有四种主要类型。第一种是持续时间，表示单个阶段或整个过程所花费的时间。无论持续时间 是否是SPS模型中明确标示的度量指标，模型都会在一段时间内模拟真实过程。日历日期 可作为持续时间 的替代度量指标，适用于模型中存在多个重叠的持续时间度量指标的情况。在用于风险管理的随机仿真模型中，会测量每项风险的开始日期和结束日期 [Zhou 12]。第三种是工作时间，常在模型中用作乘数，以模拟每天工作小时数 [Wake 05]。延迟是最后一种时间度量指标，例如招聘延迟、从完成到发布的延迟、新需求反馈延迟等。

作为SPS中制品规模的一种替代度量指标，任务规模 贯穿开发过程的各个阶段，用作相应过程（阶段）作业规模的指标。

变更 发生在需求、设计、实现阶段，或贯穿整个过程。它是导致返工的主要原因之一。根据所综述的研究，这类度量指标在SPS模型中并未得到广泛使用。

增量 表示新需求或增量加入开发过程的过程。新增需求数 和新需求生成率都是用于度量这一过程的指标。

返工 是变更 和缺陷活动的反馈。需要返工或已经返工的任务数量和百分比 是返工 度量指标。我们将返工率归入工作速率 类别，以强调过程的活动和“速度”。

我们在产品实体中引入了缺陷 。此外，还有一些与缺陷相关的度量指标应属于过程度量指标，我们将其归入缺陷活动。缺陷 度量指标衡量产品缺陷的静态属性，例如缺陷密度。

实体	属性	SUB	度量指标
	Time (300)	持续时间	间隔需求 (Days)；项目持续时间 (小时、天，周、月)；Task持续时间 (Days)；修复持续时间 (小时、天)；答案time (Days)；Lead time (Days)；交付time (Days)；需求选择时间 (天/需求)；编码持续时间 (Days)；综述持续时间 (小时、天)；测试持续时间 (小时、天，周、月)；返工时长 (小时、天，周)；问题评估持续时间 (小时)
		延迟	新需求反馈延迟 (月)；领域发现延迟 (月)；分析发现延迟 (月)；设计填充延迟 (月)
		Date	项目完成日期 (Date)；项目开始Date (Date)；检查开始时间检查 (日期时间)；风险发生时间的风险 (日期时间)
		工作时间	工作时间乘数 (小时/天、小时/周)
	Task (114)	Size	每次检查的任务规模检查 (页面数、代码行数)；测试任务Size (# 需求)；# 维护请求 (# 请求)；实施工作量 (# 任务、用户故事)；总计任务 (# 任务)；必需任务 (# 每位用户的任务数，# 用户故事)
		进度	百分比任务完成度 (数值数据 [0,1])；任务增加百分比由于需求波动性 (数值数据 [0,1])；任务返工百分比由于需求波动性 (数值数据 [0,1])
		Rate	任务率 (每项任务的人周数、每项任务的小时数)
		依赖关系	任务依赖关系 (后续任务开始前必须消耗的任务相关工作量百分比)
		级别	项目内任务低估程度 (任务低估百分比)；需重新编制预算的任务程度 (数值数据 [0,1])
	变更 (54)	Size	# 需求变化 (# 变化)；# 设计变化 (# 变化)；# 代码变化 (# 变化)；累计已发布的变更 (# 变化)
		Rate	需求变更率变更输入 (# 变更/时间)；变更请求接受率 (%)；变化率完成度 (每日变更数)；需求波动率 (%)
		分布	的百分比产品变更 (%)；产品变化百分比由于需求波动性 (%)
	增量 (22)	Size	规格说明增量 (# 规格说明文档)；高层设计增量 (# 增量)；低层设计增量 (# 增量)；代码增量 (LOC)；新功能需求 (# 功能)
		Rate	新规格生成率 (# 每日规格数)；新需求生成率 (# 每日需求数)；新需求反馈率 (# 每日需求数)
	返工 (6)	Size	返工量 (# 返工)
		百分比	返工百分率 (返工百分比)
	缺陷活动 (78)	注入率	编码期间的缺陷注入率 (缺陷数/代码行，缺陷/小时，缺陷数/日，缺陷数/任务，缺陷数/单元)；缺陷注入率由每位开发人员 (缺陷数/代码行)；缺陷再生Rate (缺陷百分比)
		检测率	缺陷检测率 (缺陷/小时)；个人缺陷检测率 (缺陷/小时)
		修复率	缺陷修复率 (缺陷/小时)
	Work Rate (117)	开发Rate	开发率 (DSI/天、行/天、FP/天、任务/迭代)
		设计率	设计率 (文档数量、文档/月)
		实现Rate	实施率 (行/天)
		验证与确认速率	测试速率 (行/天)；组件验证率 (组件/时间)；系统验证率 (元素/时间)
		检查速率	检查速率 (NA)
		返工速率	返工速率 (每个缺陷所需小时数，任务/天)；返工发现率 (任务/小时)；修订速率 (任务/天)
		发布率	交付速率 (单位/时间)；发布率 (模块/月)
	工作量 (149)	总工作量	总工作量 (人时、人日、人周、人月)；有效工作量 (人日)；计划工作量 (人日)
		单位工作量	每个测试用例的工作量 (人周)；测试工作量率 (小时/单位)；测试工作量 (人日)；会议工作量 (人时)
		活动工作量	返工工作量 (人日)；修复工作量 (人周，人日/缺陷)；准备工作量 (人日)
		阶段工作量	设计工作量 (小时)；编码工作量 (人时)；评审工作量 (人日)；测试工作量 (人日)
	开销 (17)	沟通开销	沟通开销 (数值数据)
		培训开销	培训开销 (数值数据)
		其他	任务切换开销 (分钟；小时)

图 7：过程度量分类（第1部分）

实体	属性	SUB	度量指标				
	Risk (13)	Size	# 已发生的风险 风险	已识别风险数 风险			
		影响	风险发生的影响 数值数据 [0,1]	所有已发生风险的总体影响 数值数据 [0,1]			
	过程因素 (110)	政策	结对编程策略 布尔开关	评审调整策略 布尔开关			
		方法	编码方法 由于所用方法导致的千行代码/天；# 由于所用方法导致的每千行代码缺陷数	阅读技术 序数（不同技术）			
		进度压力	进度压力 正常工作量百分比				
		发现因子	领域发现因子 已通过变更请求百分比	分析发现因子 已通过变更请求百分比	设计发现接受因子 已通过变更请求百分比		
		其他	需求波动性乘数 数值数据	新需求反馈因子 数值数据	维护调整因子 数值数据	工作量乘数 数值数据	
	人员连续性 (33)	Rate	人员流动 Rate 人日	招聘 Rate 人日	吸收 Rate 人日	学习 Rate 经验/天	Loss Rate 人日
	过程静态状态 (20)	开发阶段 开发	软件开发阶段 序数（阶段）	冲刺编号 序数（编号）			
		其他	# 检查 检查	# 过程增量 过程增量			
	Cost (52)	单项成本	单项成本 $	测试人员成本 美元/小时			
		总成本	项目总成本 $	已批准预算 $			
		活动成本	缺陷成本 每个缺陷所需小时数	单位工作量成本 美元/人日	实施成本 $	截止日期延误惩罚 美元/天	
	有效性 (35)	有效性	测试有效性 缺陷检出率	评审有效性 缺陷检出率	验证与确认过程的有效性 缺陷检出率		
	过程绩效(6)	绩效	成本绩效指数 数值数据	进度绩效指数 数值数据	开发绩效 工作量/小时		

图 8：过程度量分类（第 2 部分）

缺陷活动中的度量指标用于衡量动态活动的属性，例如编码期间的缺陷注入率。该比率可以用单位时间内注入的缺陷数量来衡量，也可以用每单位制品规模（代码行数）的缺陷数量来衡量。

工作速率表示活动执行的“速度”，涵盖整个生命周期，包括设计、实现、综述、测试、发布等过程。开发率是最常见的一种，用于衡量模型中模拟的主要活动。仅使用开发率会将整个过程视为一个整体，而不会区分过程中不同活动的速度。相反，可以分别用设计率、实施率等来衡量不同活动的速度。

工作量在许多模型中都得到应用（或研究），因为它是过程建模特别关注的内容。其度量指标例如设计工作量、返工工作量，甚至总工作量，对应于除需求阶段外的各个阶段。工作量通常以人时、人日和人月衡量，具体取决于模型的分辨率。在软件过程研究中，工作量被视为成本或其替代项 [Rus 14] 的主要贡献因素。

风险在 SPS 模型中很少出现。在旨在研究风险影响的模型中，模型使用了风险的数量、影响和发生时间 [Zhou 12]。

开销 源于无法避免在任务之外进行额外工作的情况。沟通开销是最常见的开销类型。我们将其归为过程类别，因为它产生于团队协作中的沟通活动（或其他活动）。其他类型的开销，例如任务切换开销 和训练开销 也在若干研究中得到考虑 [Baum 17, Garo 16]。

过程因素通常指管理因素和组织因素，它们会影响整个过程。在 SPS 模型中，政策是这一子类别中最常用的度量指标，但在不同模型之间差异显著。进度压力是生产率或工作速率的常见影响因素。研究中使用了各种因素，此处难以一一列举。多种发现因素用于确定可能进入下一阶段的要素比例，可利用历史数据进行校准 [Wern 02]。应考虑哪些因素，取决于具体的项目特征和组织特征。

人员连续性与培训和人员流动过程有关，而度量指标衡量这些过程的“速度”。例如，为了模拟人力资源演变过程，可以考虑招聘率、解雇率和人员流失率 [Ruiz 01]。

过程静态状态包含预先定义且在建模过程中不会改变的度量指标，例如开发阶段生命周期、# 过程增量等。开发阶段可以与系统类型等产品特征结合，用于衡量能力 [Hurt 15]。

过程外部度量偶尔会在所综述的模型中出现。它们包括有效性、成本和过程绩效，其中有效性和成本被大量使用。对于软件过程，可以将许多关注点定义为有效性指标，例如测试需求缺陷检测有效性 [Cang 08]。同样，在针对各种目的开展的 SPS 研究中，成本通常以货币价值和工作量来衡量，例如编码人员成本、估算预算、缺陷修复成本等。

6.1.5 资源度量指标



图9：资源度量指标分类

与产品和过程不同，图 9 显示，建模人员在资源属性方面更关注外部实体，因为资源本身只有规模或角色可以测量。诸如人力技能之类的资源属性，无法仅通过资源本身直接测量，而必须借助其他外部参照。

人力度量开发所需的人力资源数量。根据建模粒度和模拟项目的规模，研究可能将所有资源视为一个团队，也可能将个人分配到不同阶段。

环境很少有研究在SPS模型中考虑环境。例如，测试设施可用性被引入，用于模拟资源对测试阶段的限制 [Hous 10]。项目是否属于多地点开发，多地点开发是COCOMO II提出的一个因素。

为了量化人力在过程中的作用或价值，可以测量五项基本的人力属性，即生产率，能力、专业水平、经验和技能 。生产率 构成了开发率的基础。其余四项度量指标是决定生产率的因素。其余四项度量指标是决定生产率的因素。专业水平 和知识 是同一项度量指标。与专业水平不同，经验 已经考虑了领域知识，而且往往只用年数来衡量，缺乏历史数据。

人力因子包含所有与资源相关、具有高度多样性的影响因素度量指标，范围从自尊、团队凝聚力到母语。

现有研究仅考虑人力资源利用率 ，虽然研究中使用了个人和团队利用率，但这只出现在少数研究中。

研究发现： 1）大多数综述研究只详细描述被认为重要的度量指标。2）在过程仿真建模中，产品和过程外部度量使用不频繁，而资源外部度量则被广泛使用。

6.2 度量之间的因果关系（研究问题2）

SPS模型旨在仿真开发团队在环境的约束与支持下，通过开展一系列开发相关活动来生产软件产品的过程。在这一过程中，各项活动会按照特定工作流组织起来并相互影响，而人是活动的具体参与者。人会受到环境及各种其他因素的影响，也可能对这些因素作出反应。SPS模型使用不同的模块（由仿真范式定义）来描绘过程中的不同要素，包括人员、活动、环境等。这些模块是在特定仿真范式下对本研究所讨论度量指标的实例化实现。在实现层面，SPS模型由模块及其相互关系构成。模型的实现描绘了现实中的不同度量指标及其相互关系。我们在研究问题1中讨论了现有研究使用的度量指标，研究问题2将讨论度量指标之间的因果关系。

我们收集了纳入论文中所有可识别的因果关系（从度量指标 A 到度量指标 B）。最终，我们识别出 183 种因果关系。表7列出了高频因果关系（在所有 SPS 模型中出现超过 10 次；同一关系可能在一个模型中出现多次），我们将在下文对此进行更详细的讨论。

从人力属性到人力属性。这是最常用的因果关系。软件开发人员与软件开发活动之间最直接的关系是度量指标生产率，它描述了个人参与活动的能力。生产率是人力属性度量指标中最常见的一项。生产率会受到经验和知识等因素的影响，这些也属于人力属性 [Hana 98、Klun 18、Oors 18]。此外，生产率还可以进一步细分。例如，实时生产率可由增长和基线/平均生产率构成 [Lehm 10、Zawe 13]。诸如经验这样的度量指标也可以进一步细分。例如，检查人员的经验可以

受到开发经验、检查经验和领域经验 [Neu 02b]。

从缺陷规模到缺陷规模。缺陷在软件生命周期中可能处于不同状态。状态的变化会反映在数量的变化上。缺陷总计可以是未解决缺陷、进行中的缺陷、待测试缺陷、重新打开的缺陷和已解决并关闭的缺陷之和 [Zhan 18]。此外，已检测到的缺陷数量会受到生成的缺陷数量的影响 [Zhan 08a]。逃逸缺陷数量可以是已检测到的缺陷数量与已修复的缺陷数量之差 [Lake 03]。此外，缺陷也可以按不同类型描述。已检测到的缺陷总计可以是已检测到的被动缺陷和已检测到的主动缺陷之和 [Zhan 08b]。

从变更到变更。这种类型的因果关系主要见于研究 [Zhan 18] 和研究 [Klun 18]。在研究 [Klun 18], 中，迭代变更能力基于每日变更能力。其余关系见于研究 [Zhan 18], ，其中变更也称为工作中的问题请求。变更基于问题跟踪系统中的问题生命周期进行了详细建模。例如，待审查问题数量会受到迭代问题数量的影响。问题总数是重复和无效问题、未解决问题、进行中的问题、迭代问题、待审查问题、已解决并关闭的问题和重新打开的问题之和。

从工作速率到工作速率。一种工作速率可以由许多其他速率合成。例如，名义开发速率可以是有经验员工的开发速率与新员工开发速率之和 [Zhan 06b]。此外，任何类型的实际工作速率都可以基于某种基准工作速率 [Zhan 18]。

从制品规模到制品规模。制品可能有不同的来源，或处于开发的不同阶段。项目规模可以是剩余规模与已完成规模之和 [Zhan 06b]。需求生成率可能受到生成的新需求数量反馈的影响 [Zhan 09]。生成的新需求数量可以基于外生需求的数量 [Hall 05]。

从制品属性到制品属性。制品属性的细化可能不太常见，但可以非常细粒度。研究 [AlE 08] 使用了 15 种不同的复杂度度量指标来量化软件系统的综合复杂度。研究 [Aran 08] 对以下指标之间的关系进行了建模：骨架中识别出的控制流数量、用例中编写且在检查后需要返工的控制流数量，以及用例中编写的控制流数量。研究 [Aker 17] 使用测试代码的歧义性和脆弱性来量化测试代码异味。

起点过程因子终点过程因子。一个过程因子可以分解为若干不同的因素。例如，分解相关性基于多个因素的组合，包括相关生产率、冲刺周期相关性、评审相关性、承诺与忠诚度相关性以及客户满意度相关性。

从 时间到 时间。时间通常用于描述某项活动的持续时间或延迟。一项活动包含开始时间和结束时间，两者之间的差值就是持续时间 [Zhou 12]。某项活动的持续时间或延迟可以基于平均值或基线值 [Ruiz02b]。平均生产时间可能会受到工作分区导致的时间损失影响 [Hsia 99]。

起点人力终点人力。人力可根据经验或其他因素进行区分。员工总数可以由以下两项相加得出：受训员工数量和有经验的员工数量 [Zawe 13]。有经验的员工可以由以下两项相加得出：开发方面的有经验员工和培训方面的有经验员工 *Zhang2006Semi*。每日可用劳动力可以根据总劳动力 [Ambr 11]。

从工作速率到制品规模。这类关系描述了一种常见情况：制品规模会根据工作速率累积。例如，待处理的规格单元取决于规格单元完成率 [Wern 99]。此外，将软件开发速率与其他因素结合，也可以决定哪些需求未得到正确满足。在这类关系中，后者并非前者的简单累积 [Hall 05]。

从工作量到工作量。工作量可根据不同活动加以区分。例如，集成工作量可以是项目评估工作量、项目裁剪工作量和胶水代码工作量之和 [Naun 07]。此外，开发工作量可根据预期开发工作量和增长工作量计算 [Choi 06]。

从环境到人力属性。环境是影响人力的主要因素。例如，市场薪酬、工作环境、团队管理和奖励等环境因素可能会影响开发人员的动机 [Fate 18]。

从人力属性到工作速率。开发速率（工作速率）取决于人类生产率（人力属性）。例如，新员工开发速率取决于新聘员工的工作效率 [Zhan 06b]。

研究发现：1）高频因果关系主要是同类型度量指标之间的关系。这表明，SPS模型通常会细化主要度量指标类型，例如人力属性、缺陷规模、工作速率、变更、制品规模、制品属性、过程因子、时间、人力等。2）此外，尽管软件生命周期的主线十分清晰，但我们没有发现任何一种关系出现在超过10%的模型中。这充分体现了SPS模型的多样性：针对具体问题需要进行具体分析，而且即使是同一个问题，也可能采用不同的实现方式。

6.3 基于目的和范式的指标选择（研究问题3）

如元模型所示，用于构建仿真模型的度量指标取决于构建相应概念模型的信息需求，而信息需求又取决于建模目的。Kellner 等人 [1]也指出，要仿真的许多方面彼此相关，并由目的驱动。建模人员会以不同的粒度级别仿真过程，因为建模人员几乎不可能、且在某些情况下也没有必要仿真过程的每个细节。因此，

表7：度量指标之间的高频因果关系

因果关系		频率	引文
From	**To**		
人力属性	人力属性	44	[Mada07、Lehm10、Zawe13、Neu02b、Neu02a、Zhan08a、Cher 10, Spas 12、Oors 18、Hana 98、Fate 18、Klun 18]
缺陷规模	缺陷规模	22	[Lake 03, Mart 00, Zhan 08a, Zhan 08b, Zhan 18]
变更	变更	22	[Zhan 18, Klun 18]
工作速率	工作速率	20	[Zhan 06b、Wern 02、Oors 18、Wern 99、Zhan 18]
制品规模	制品规模	19	[Zhan 06b, Hall 05, Aran 08, Zhan 09, Wern 99, Zhan 18, Klun 18, Duga 21]
制品属性	制品属性	18	[Aran 08, Shuk 12, Aker 17]
过程因子	过程因子	15	[Oors 18、Zhan 18、Klun 18]
Time	Time	15	[Hsia 99、Zhou 12、Mish 16、Wake 04、Ruiz 02b、Zhan 18、Klun 18]
人力	人力	12	[Zawe 13、Mada 05、Zhan 06b、Ambr 11、Tram 16、Kahe 01、Zhan 18]
工作速率	制品规模	12	[Lehm 10, Wern 02, Hall 05, Zhan 09, Wern 99, Zhan 18, Duga 21]
工作量	工作量	10	[Choi 06、Naun 07、Aran 08、Noll 16]
环境	人力属性	10	[Hurt 15、Fate 18、Klun 18]
人力属性	工作速率	10	[Lehm 10, Zhan 06b, Mada 00, Wern 02, Zhan 08a, Cocc 11, Mish 16, Oors 18]

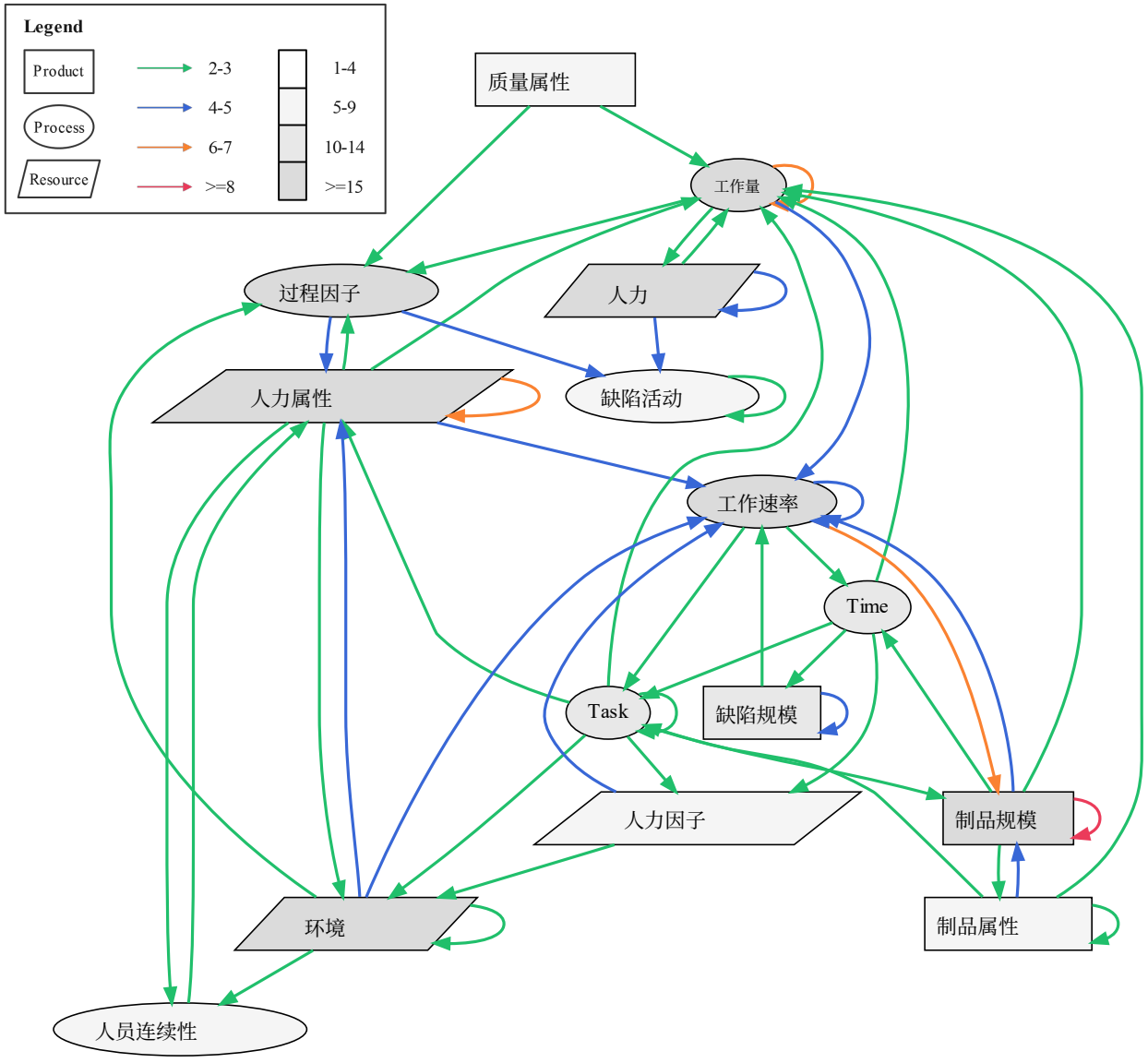

Fig. 10：认知模型中度量指标之间因果关系的图示（频率 ≥ 2 ）.

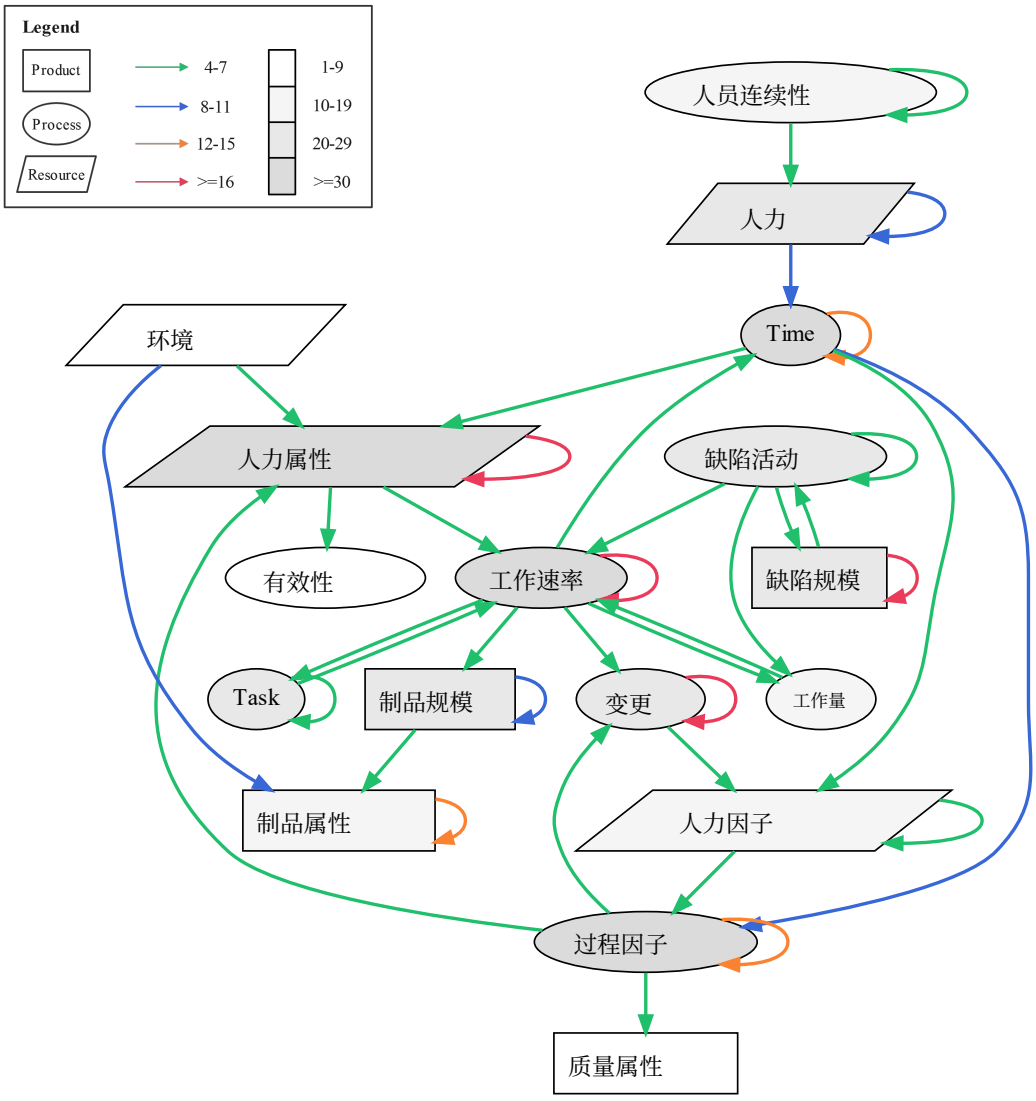

图11：因果图 图 战术与 战略模型中的关系（频率 ≥ 4）。

筛选适当的度量指标对于满足不同建模粒度的需求至关重要。在很大程度上，这取决于研究的具体目的。如图5所示，根据建模目的，现有的SPS模型被分为认知模型（CMs）和战术与战略模型（TSMs）。研究问题3从所使用的因果关系这一角度讨论了CMs与TSMs的共性和差异。

此外，软件过程的多样性和复杂性（可体现在建模目的中）决定了所需仿真范式的不同能力 [25]。研究问题3还将讨论不同范式之间建模粒度的差异。

6.3.1 建模中两个目的级别的比较

我们在 CMs 中识别出 217 个因果关系（根据频率，同一类型可能被重复计数），在 TSMs 中识别出 462 个因果关系。图10和图11分别展示了 CMs 和 TSMs 中度量指标之间因果关系的示意图。如果纳入所有已识别的关系，示意图将过于复杂，难以理解。经过权衡，我们最终将频率大于或等于 2 次（约占总数的 1%）设为 CMs 的阈值。考虑到在 TSMs 中识别出的关系数量约为 CMs 中的两倍，我们将 TSMs 的阈值设为 4 次。因此，图10共包含 51 种因果关系（同一类型仅计数 1 次），涉及 15 种度量指标；图11共包含 39 种因果关系（同一类型仅计数 1 次），涉及 17 种度量指标。

在图示中，我们分别使用矩形、椭圆和菱形表示产品、过程和资源度量指标。填充颜色越深，表示该度量指标的使用频率越高。我们使用不同颜色的箭头曲线表示关系的频率。绿色、蓝色、

橙色和红色分别表示 CMs（TSMs）的频率大于或等于 2（4）、4（8）和 8（16）。

从度量指标的角度来看：CMs 和 TMs 中高频度量指标存在明显重叠，包括制品规模、缺陷规模、工作速率、任务、时间、过程因子、人力和人力属性。这些都代表软件过程的基本要素，即参与者、制品和活动。借助这些度量指标，可以从宏观层面描述软件开发过程。这或许是它们出现频率较高的原因。TSMs 中工作量 的出现频率不高，但在 CMs 中较高。TSMs 中变更 和缺陷活动 的出现频率较高，但在 CMs 中不高。此外，TSMs 中有有效性 ，而 CMs 中没有。

从因果关系的角度来看：CMs 和 TMs 中都有四类高频关系，包括从制品规模到制品规模、从缺陷规模到缺陷规模、从人力到人力以及从人力属性到人力属性。这表明，CMs 和 TMs 在制品、人力和缺陷的建模细致程度方面存在共性，也就是说，对不同制品、人力和缺陷的区分，以及对人力属性的区分，都得到了进一步细化。需要指出的是，尽管从人力属性到人力属性在 CMs 和 TSMs 中都是高频关系，但其在 TSMs 中的频率高达 38（8.2%），而在 CMs 中仅为 6（2.8%）。此外，TSMs 和 CMs 中从缺陷规模到缺陷规模 的频率分别为 18（3.9%）和 4（1.8%）。CMs 中大多数高频关系的形式是从某个度量指标到工作速率，而这类关系在 TSMs 中并不属于高频关系（从工作速率到工作速率除外）。即便如此，在 CMs 和 TSMs 中，工作速率 周围的关系类型都是最多样的。从时间到过程因子 和从人力到时间 是 TSMs 中的高频关系，而 CMs 中没有这类关系。从工作量到工作量 在 CMs 中是高频关系，但在 TSMs 中不是。从制品属性到制品属性、从时间到时间和从过程因子到过程因子 在 TSMs 中是高频关系，但在 CMs 中不是。

研究发现： 在 TSMs 中，同类型度量指标之间的因果关系明显更多；而 CMs 中围绕工作速率、制品规模和工作量的关系则更常见。这表明，CMs 倾向于在宏观粒度上直接围绕最终结果（例如工作量、输出制品）建模，而 TSMs 则更关注微观粒度。TSMs 倾向于在同一类型中使用多个度量指标，对某个方面（例如过程因子、人力属性、缺陷规模等）进行详细建模。

6.3.2 不同建模范式在建模中的比较

为了仿真不同粒度的过程，需要采用适当的建模范式。例如，ABS 能够将每位开发人员仿真为一个主体，以研究模块复杂性对个人生产率和动机的影响。在这种精细粒度级别上，由于需要分别仿真不同的开发人员，使用系统动力学范式构建的模型会非常复杂。在现有研究中，DES 和 ABS 范式从未用于仅停留在认知层级的模型。采用不同的仿真范式意味着某些度量指标也会发生变化。对于不同的仿真范式，用于衡量同一属性的度量指标需要随之改变，这在一定程度上由建模粒度决定。下面列举一些示例来说明主要差异。

仅在 DES 和统计模型中使用表示事件或状态可能发生变化概率的概率度量指标。例如，DES模型使用基础概率 $P(D_k^i(t))$ 来指定团队 i在投入 t 个时间单位 [45]后完成组件 k 的可能性。如果是在系统动力学模型中，则可以使用每天完成的组件数 来表示工作速率。对于擅长处理宏观过程的范式（如系统动力学）而言，模拟不同团队将不同组件分别作为独立事件实施的活动会很复杂。

出于类似原因，ABS 模型更擅长详细建模不同的开发人员。例如，Cherif 和 Davidsson [46]构建了一个 ABS 模型，用于模拟任务 j 的难度，即知识水平之间的差异b_{ij} 以及所需的知识水平 θ_j，其中 b_{ij} 表示开发人员 i 对活动的知识 j。Cherif 和 Davidsson 将该 ABS 模型与系统动力学模型进行了比较，并指出二者的主要区别在于，系统动力学模型将个体特征（人力属性）的平均值作为替代值输入。

研究结果： 与系统动力学模型相比，DES 和 ABS 分别能够从开发活动和个人开发人员的角度提供更详细的仿真，这体现在它们采用的度量指标更为详细。相比之下，系统动力学倾向于使用平均值作为替代值。这样既降低了获取度量指标的难度，又能减少对宏观过程研究的影响。

6.4 研究问题4：重新审视数据问题及应对策略

早在二十年前，Raffo 和 Kellner 等人 [1, 39]就讨论了在仿真中测量过程度量指标时可能出现的若干情况及相应的应对策略，为后续 SPS模型研究奠定了基础。多年来，SPS模型面临着许多新的挑战。数据问题无疑是 SPS 面临的主要挑战之一。

在对纳入的145篇论文进行综述后，我们通过检索总结了仿真建模数据准备阶段遇到的大多数数据问题，尽管并非每篇论文都通过建模充分讨论了自身的数据问题。此外，我们将两项相关研究 [39, 47] 纳入，作为研究问题4的补充数据来源。我们在文献筛选阶段发现了这两项研究。尽管这两项研究不符合筛选标准C1和C2，但它们报告了与数据问题相关的经验。纳入这两项研究有助于我们更全面地讨论数据问题。

表8：SPS数据准备过程中遇到的数据问题（来自我们纳入的研究和相关工作的证据）。

Data 准备步骤	遇到的数据问题	具体问题	度量指标（例如，）	Ref.
数据溯源	可用性	实际项目中没有现成的元数据 模型中的变量需要使用文献中未提供的数据 文献 记录缺失 无可用的定量信息	返工工作量 需求波动性；受影响 作业规模（百分比） 未指定 测试驱动开发如何减少缺陷 被拒绝的数量	[Ruiz 01，Ruiz 02a， Neu 02b、Berl 03、Ruiz 04， Choi 06、Chen 06、Turn 06， Park 07、Topi 08、Ferr 09， Dan 13、Baum 17], [39]
	稀缺性	使用的样本仅限于少数几个项目 样本可用信息的稀缺性	未指定 #维护请求	[Turn 06，Conc 13， Hurt 15], [39]
	可靠性 元数据	存储库数据中可能存在系统性偏差 由于缺乏对 一些参数	任务状态 未指定	[Kous 07, Baum 17],[47]
	多样性	因组织结构不同而导致的数据多样性 和项目环境	未指定	[47]
	可追溯性	不同制品和来源之间的数据“孤岛”	需求；缺陷	[47]
数据收集&处理	理解	对数据真实含义的误解 难以找到数据字段的准确描述 因为分散在各个数据来源中	未指定 未指定	[47]
	可访问性	各种接口的限制	未指定	[47]
	时间跨度	难以选择历史数据的时间跨度	未指定	[47],[浓度 13]
	处理	时间和工作量消耗 存在相当大的变异、离群值和噪声	软件开发人员时间与项目 成本与工作量 未指定	[Pfah 99, Raff 00, Ferr 09] [Conc 13，Lune 17], [39]
数据测量	定义	定义宽泛	未指定	[39]
	粒度	仿真模型是对现实世界过程的抽象， 度量指标比实际数据记录更粗略。	未指定	[Raff 03, Al E 08, Zhan 12b]
	准确性	模型的许多参数只能估算， 测量结果的准确性存疑	未指定	[Baum 17], [39]
	完整性	无法准确测量所有变量及其相关性 准确地	未指定	[Choi 06],[47]
	测量方法	现有研究依赖于使用工业数据平 均值、专家估算，或从分析 模型	未指定	[39]
	可量化性	不可量化变量 不可量化变量的关系	技能水平、人力、沟通 开销、评审支持， 流程成熟度和工具支 持值 交互 之间 开发人员->维护工作量	[Choi 06, Armb 03, Conc 13, Lune 17] [Choi 06、Xiao 10，Conc 13]
	动态	不了解变量的经验分布 变化	开发人员的增长和变化 行为；工作量变化；开发 人员的分配	[Baum 17，Lune 21]
	结果可靠性 结果	上述度量是否能提供可靠的	未指定	[Choi 06、Klun 18、Lune 21]

6.4.1 SPS 数据准备中遇到的数据问题

具体而言，我们重点关注构建SPS模型时遇到的数据问题。为便于理解，我们按照SPS数据准备过程将数据问题划分为三个步骤，即数据来源问题、数据收集与处理问题以及数据测量问题。

数据溯源问题。

1) 可用性。软件过程会产生各种复杂数据，而数据可用性一直是仿真建模研究面临的最大挑战。全部 17 项研究都表明，数据可用性是一个问题。并非所有活动都有文档记录，而且大多数数据无法从相关文献中获取。存在缺失数据记录，例如，没有任何活动的返工工作量记录 [Choi 06]。尤其是，无法获取定量数据来源（例如，需求波动性导致的工作规模增加比率 [Ferr 09]）。

2) 稀缺性。即使能够获取数据，数据稀缺也是建模人员普遍面临的问题。在大多数仿真研究中，可用于仿真的数据寥寥无几，因为选定项目的历史记录太短 [Conc 13]。项目记录数据普遍存在的问题是，无法提供足够的信息来仿真真实的变量变化及其相互关系 [Turn 06]。

3) 元数据可靠性。 建模人员接触真实的数据来源时，会关注数据的真实性和可靠性。定量和定性数据均来自软件存储库、手工记录文档和专家估计，而这些数据来源或多或少都容易出错。例如，软件开发人员可能会错误地选择缺陷类型、忘记缺陷定位时间、延迟更新需求和进度计划文档 [Baum 17],[47]。

4) 多样性。由于数据类型、记录形式、项目流程和组织

结构各不相同，过程数据自然也会呈现多种形式[Hurt 15]。由于组织和项目开发环境不同，数据来源可能分布于各种线上和线下来源之间，彼此之间没有直接相关性。如果仿真模型涉及软件过程多个阶段和活动的数据，就很难确保数据的一致性。

5）可追溯性。通过建立过程数据之间的可追溯性，还原其真实演变过程具有重要意义。以往研究很少讨论这一问题，也未说明如何处理这一问题。Ali 等人 [47]指出，实际项目中的过程数据信息存在“孤岛”。需求存储库保存需求相关的过程信息，而缺陷报告则存储在另一个数据库，即问题跟踪系统中。因此，尽管这两类数据都存在，但由于它们之间的关联缺失，我们无法利用这些数据。

数据收集与处理问题。

即使建模所需的数据来源已经具备，建模人员仍需面对如何收集数据，以及如何处理数据以满足后续仿真建模需求的问题。

1）可理解性。过程数据可能分散在各部门的不同数据来源中，因此可能难以理解文档中数据字段描述的准确含义 [47]。即使字段名称看似直观，建模人员也必须经常向领域专家咨询数据字段的含义。不同组织、团队和数据库中的字段名称各不相同。例如，在问题跟踪系统中，发布版本号可能被定义为“基线”，而在需求管理存储库中则可能被定义为“版本”。

2) 可访问性。为了从多个数据来源导出数据，必须能够访问各种接口。访问这些接口仍需获得各业务部门的权限和授权，因此数据收集过程极其复杂且耗时。由于数据敏感性和权限方面的顾虑，数据收集可能会失败。

3) 时间跨度。Ali等人 [47]建议与领域专家讨论构建和校准SPS模型所需历史数据的时间跨度。为了模拟变量的变化分布及其相互作用，需要足够长的时间段 [Conc 13]。使用历史数据模拟当前现实情况时，需要有足够的历史数据，以便充分了解潜在变化。例如，如果开发过程、编程语言或开发平台发生变化，建模人员就需要了解这些变化。

4) 处理。 获取、收集、处理和分析数据无疑是一个复杂的过程，需要投入大量时间和工作量 [Pfah 99, Ferr 09, Raff 00] ，因为其中涉及数据多样性、可靠性、可理解性等问题。目前，SPS研究中的数据收集和处理活动仍依赖手动方法，例如数据格式转换、离群值处理和噪声数据处理。

数据测量问题。

数据准备的主要目的应是提供仿真模型测量模型度量指标所需的数据，这需要测量模型变量的实际值及其关系。因此，SPS非常重视度量指标的测量。根据对先前工作中报告的测量问题的总结，我们将其分为八类，即定义、粒度、准确性、完整性、测量方法、可量化性、度量指标的动态性，以及SPS模型结果的可靠性。

1) 定义。元模型表明，SPS模型中使用的度量指标基于构建概念模型时采用的信息需求，而信息需求取决于建模目的。由于认知差异，建模人员可能无法理解开发过程的细节。他们可能会忽略测量的难度，因而定义出不恰当的度量指标。

2) 粒度。 建模人员通常无法模拟过程中的每个细节，因为这样做要么不可能，要么没有必要。因此，应根据建模粒度来选择和测量度量指标。然而，在许多情况下，现实世界数据的粒度过细，无法与模型的假设相匹配 [Zhan 12b]。例如，系统动力学模型假设输入变量（如新的需求）之间没有差异。然而，实际情况下，不同需求之间存在巨大差异，包括难度、优先级、依赖关系，以及应投入其中的开发人员数量 [Al E 08]。这两个因素都可能影响工作规模。

3) 准确性。 测量的准确性也是SPS建模中的常见问题。由于数据质量和测量方法的选择，准确性常受到质疑。当可用数据来源有限时，模型中的许多变量不得不由专家估算，或参考其他文献。然而，专家估算源自主观认知，且始终受限于直觉不足 [Baum17,Klun 18]。

4) 完整性。 此外，并非所有变量都能得到准确的定义和测量。由于数据可用性、质量和度量测量方面的问题，建模人员不得不调整模型的结构和范围，这也可能影响SPS模型的完整性。

5) 测量方法。 测量方法通常包括均值估计、概率估计、分布拟合，以及使用蒙特卡洛进行随机模拟。此外，现有SPS研究采用行业平均值、专家经验估算，或分析模型推导出的数值 [48]。这些方法过于简单粗略，无法满足建模需求，难以捕捉变量的实际变化及其相互影响。

6）可量化性。许多过程度量指标，例如人的方面因素 [Choi 06] （如技能水平、开发者之间的交互），都难以量化，但对仿真模型至关重要，因为它们会极大地影响生产率 [Lune 17]。准确量化变量之间的因果关系是一项经典的挑战性任务[47]。

7）动态。尽管数据来源和质量有所改善，建模人员仍然无法根据历史数据拟合变量的真实动态分布。例如，

编号	应对策略	Ref.
CS1	直接查阅源文档和代码库。	[Baum 17, Raff 00, Menz 02]
CS2	查看组织其他部门是否有数据	[Menz 02],[47]
CS3	制定一份调查问卷以收集所需数据。	[Menz 02, Ferr 09]
CS4	与有经验的流程参与者合作或征询专家意见，例如开发人员、项目经理。	[Pfah 99, Menz 02, Chen 06, Topi 08, Rus 14], [39]
CS5	从文献获取数据。	[Berl 03, Ruiz 04, Turn 06, Rus 14]
CS6	使用近似数据或易于配置的数据。	[Kous 07]
CS7	调整模型变量：另一种计算方法，使用替代指标，调整变量范围。	[Baum 17], [39]
CS8	调整模型范围。	[39]
CS9	调整历史数据的尺度。	[Conc 13]
CS10	从模型中删除该变量。	[39]

很难衡量开发人员的成长，以及这种成长对其行为的影响 [Lune 21]。

8) **结果可靠性。**所有数据准备工作最终都是为了构建和校准仿真模型。这反过来会直接影响仿真结果的可靠性。换句话说，要验证并确保 SPS 模型中所建模的制品、活动和参与者足够可靠，从而能够生成合理的结果，是一项艰巨的挑战 [Choi 06, Lune 21]。为了实现建模目的，我们最终需要解决这一问题。

研究发现：1) 在建模工作的整个演进过程中，数据的可用性一直是一个主要问题。2) 此外，模型度量指标的建模需求（即对实际过程变量的变化及其相互影响关系进行建模）给建模人员带来了巨大挑战。3) 在准备过程的各个阶段都可能遇到具体问题，包括可靠性、多样性、可追溯性、历史数据的时间跨度、数据接口、数据格式和处理时间等方面的问题。

6.4.2 数据问题与应对策略之间的关系

我们确定了十种应对策略（CS1-CS10），如表9所示，其中大多数已由Raffo和Kellner讨论过。 [39], CS3和CS9是新发现的策略。图12展示了一张可视化图，用于评估数据问题、应对这些问题所采用的策略，以及SPS研究中为支持这些结论而提供的证据数量之间的关系。

多年来，数据可用性一直是建模人员面临的一项重大挑战，因此研究人员提出了各种解决方案。如果建模人员能够访问多个数据来源，就可以同时利用源文档、代码库，与有经验的从业者合作、向专家获取估算值、从文献中获取数据等，以减少偏差。表10展示了可用数据不足以支持建模时可以使用的三种常见数据来源：经验丰富的专家、文献和组织文档。

经验丰富的专家（***CS4***）是收集和验证度量指标的主要数据来源之一。例如，分布式支持部门的一位经理正在开展流程改进计划，并向我们提供了生产率可能提升幅度的估算值范围 [Pfah 04]。此外，对于定性数据，专家的建议尤为重要。

文献（***CS5***）。在某些情况下，如果无法获得项目数据，最初可以使用文献中的统计数据、度量指标和典型函数。质量保证模型中涵盖的很大一部分因果关系得到了研究 [51]，所提供数据的定量支持，该研究提供了有关错误、成本和持续时间的数据。在研究 [49]，中，分析了不同集成起点对交付时间、生产率和成本的影响。为了校准 GENSIM 2.0，Frost 等人 [52] 提供了一个缺陷遏制矩阵示例，可据此推导校准所需的故障注入率和验证有效性。Wagner [53]提供了大量有关各种文档类型缺陷的典型（平均）验证与确认率以及返工工作量的数据。

组织文档（***CS1***、***CS2***）。组织会维护各种各样的文档，其中包含许多度量指标。不同组织采用不同的过程并使用不同的工具，因此，现有证据无法涵盖所有文档类型。

如果无法获取测量数据来源，建议使用近似值（CS6）或调整变量的测量方法（CS7）。如果上述措施均告失败，建模人员可能需要再次简化模型的范围和结构，或者干脆舍弃这些变量。

研究[Pfah 99、Menz 02、Chen 06、Topi 08、Rus 14], [39]指出，建模人员会与有经验的从业者密切合作，充分理解软件过程，从而基于对数据的深入理解来构建模型。应咨询内部专家，确定项目中是否有一定规模的历史数据，以及是否已记录必要的信息，以避免建模过程中出现数据稀缺问题。

为避免因信息不对称和认知差异而重复进行数据收集与处理，研究人员建议将软件代码库、数据文档和内部从业者的意见整合起来。若不采取这些策略，就会浪费人力和时间。尽管Ali等人 [47] 提出了误解、数据可访问性风险或历史数据时间跨度等问题，相关讨论仍然很少，其他建模人员也可能面临类似挑战。

研究人员提出的数据测量新策略并不多。除此之外，我们还观察到，测量问题仍面临诸多困难，例如匹配建模粒度、量化变量模型、为变量拟合动态分布等，而很少有论文提出恰当的应对措施。毫无疑问，这将导致建模瓶颈。因此，后续仿真工作可能会受到更高的标准要求。后续研究预计将考察

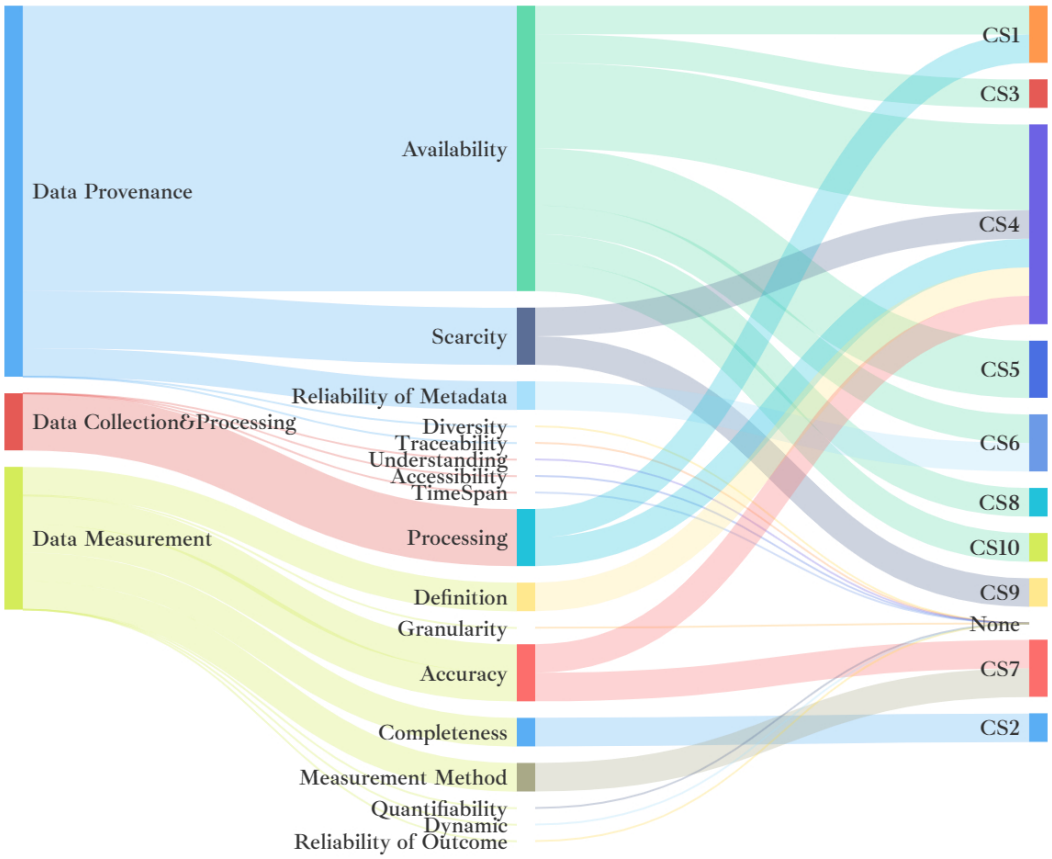

图12：数据问题与应对策略之间的关系

这些问题的更多细节，并提供全面而详细的解决方案，从而为建模人员提供有益的指导和实践参考。

研究发现：1）我们发现，从数据溯源的角度来看，讨论如何恰当应对特定数据质量问题的SPS研究相对较少，例如元数据可靠性、过程数据多样性和数据可追溯性等。2）此外，关于如何与实践者有效沟通、将数据可访问性风险降至最低，以及如何选择项目历史信息的适当时间跨度，相关细节也不够充分。3）在数据测量方面，研究人员尚未提出许多新颖的策略。然而，在现实项目中构建和应用SPS模型时，解决这些问题确实至关重要。预期后续研究能够针对这些问题提供全面而详细的解决方案。

7 有效性威胁

我们依据Zhou等人 [54]提供的映射，从四个方面检验有效性。

建构效度：找到所有相关论文是所有系统文献综述研究普遍面临的威胁。在论文筛选过程中，我们可能遗漏了少数与 SPS 相关的研究。为克服这一问题，我们对 SPS 进行了广泛研究并且检索字符串已在先前的研究中得到验证。为检索 2015 年之后发表的论文，我们还采用了 Wohlin 的前向滚雪球法策略 [55]。

内部效度： 选择偏差也是所有系统性文献综述面临的常见威胁。我们采用基于 QGS 方法 [56] 和滚雪球法的筛选方法，以减少偏差。我们使用三篇开创性论文进行前向滚雪球，这些论文被 SPS 研究中引用次数较高且最相关的研究广泛引用，从而将相关研究的范围扩展至 2021年。另一个威胁是如何尽可能减少数据提取中的错误。研究筛选和数据提取主要由两名学生独立完成，并在综合数据前进行交叉核查。导师和学生每周都会开会，讨论所有不一致和分歧。一项威胁是，大多数研究没有提供其所用度量指标的完整清单，更没有详细说明某项度量指标是否经过校准。我们只提取了研究中报告的、作者认为重要且具有普遍性的度量指标及其关系。

外部效度：我们确保所选主要研究具有较强的普遍性，这可能使研究结论也具有较强的普遍性。

结论效度：在系统文献综述（SLR）的数据综合阶段，对提取出的度量指标和数据问题进行分组是另一项挑战。我们依据Fenton等人对度量指标的分类 [22]

表10：可用数据源

数据来源	度量指标	Ref.
经验丰富的专家		
经验丰富的程序员和测试人员	待处理工作量；程序员人数；每天注入缺陷数每位程序员，等等。	[Berl 03]
工业项目经理的直觉和经验	活动的工作量；开发人员的能力；开发周期	[Hana 02]
分布式支持部门的一位经理	生产力改进的潜在范围	[Pfah 04]
对不同领域专业人士的调查	工作量	[Shuk 12]
以往的组织数据和专家判断	估算的需求波动性	[Mada 00]
专家判断	评审技能；全局阻塞问题风险；全局阻塞问题暂停时间；评审备注修复时长；问题评估时长；规划持续时间；任务切换开销；评审修复到任务的因子，等等。	[Baum 17]
文献		
Abts 等人 [49]	交付时间；生产率；成本	[Ruiz 04]
Lum 等人 [50]	*cost*	[Uzza 13a]
Jones 等人 [51]	缺陷数量;成本;质量保证模型涵盖的持续时间	[Drap 99]
Frost 等人 [52]	缺陷注入率；验证有效性	[Garo 09]
Wagner 等人 [53]	典型（平均）验证与确认率；返工工作量，针对各种文档类型中的缺陷	[Garo 09]
组织文档		
项目管理文档	进度；工作量；代码规模	[Pfah 00a, Menz 02, Raff 02]
个人检查报告	缺陷检测率；检查有效性	[Menz 02，Raff 02]
Holon 历史记录	源代码；变更历史数据；在制品估算	[聊天 00]
公司的工单系统	缺陷注入率；任务持续时间；评审持续时间；问题修复开销；# 开发人员;依赖关系	[Baum 17]
工时报告系统	新开发与返工的比率	[Pfah 04]
配置管理系统	移交生产统计数据	[Pfah 04]
IT 招聘与培训部门	人员流失率；培训统计数据	[Pfah 04]

将度量指标分为三类，并严格遵循各类别的定义，以确保度量分类的正确性。为构建一套系统的研究发现，我们在迭代过程中采用了主题综合法 [42]和编码技术 [43] 。这使我们能够根据度量指标的不同描述，形成一套一致的代码。通过应用主题综合法 [42],，我们总结了构建SPS模型时遇到的数据问题，并将其与SPS的数据准备过程相对应。

我们还面临一个常见的威胁，即原始研究缺乏详细信息。大多数原始研究都未提供此类细节，因为它们通常侧重于构建模型并分析结果。因此，我们无法全面、完整地提取和综合这些信息。由此可见，后续关于 SPS 的研究有望更详细地考察更多问题，并提供有关 SPS 度量指标的全面、详细信息。这包括测量方法、数值、类别、数据问题、应对策略和数据来源等。这可为建模人员提供有益的指南和实践参考。

8 结论

SPS模型建模人员承担着选择合适度量指标所带来的高昂成本，而社区认为这正是实践中采用SPS的关键障碍之一。本文报告了一项关于SPS模型中所用度量指标及其相关属性的系统文献综述。鉴于静态模型与动态模型之间的连续性和相似性，以及动态模型所要求的更高标准，这项工作的意义也可扩展至一般的软件过程建模。本研究以度量本体的元模型为指导，使我们能够从建模的角度全面研究SPS模型中的度量指标。

经过全面检索，我们发现截至2021年共有145篇论文报告了SPS模型。我们从这些模型中识别出2130个度量，并从测量角度将度量及其相应属性划分为六类实体类别。我们采用编码技术构建分类框架。与现有的软件测量框架相比，该框架更适合作为建模参考，因为它涵盖了SPS领域内的所有度量，并通过多层级分类展示各度量之间的相似性和差异。与Fenton等人的框架不同， [22], 我们将缺陷活动与缺陷规模及缺陷属性区分开来，并将缺陷规模和缺陷属性归入产品类别，因为这种分类更为合理。范式的选择同时受到建模粒度和可用数据的制约。因此，我们进一步讨论了测量中的数据问题及应对策略。

本研究有助于应对建模中的两项主要挑战性任务之一，即从实际过程中识别关键度量指标（模型的要素）及其关系（模型的结构）。尽管我们讨论了现有研究报告的数据问题和应对策略，但对于建模人员而言，从工业环境中收集数据仍是一项艰巨的任务。SPS研究通常假设组织能够收集所需信息，但人们也认识到，在某些情况下这并不可行。近年来，过程挖掘已成为从各种软件存储库中提取过程信息的有效工具。我们在此前的研究 [57] 中归纳了许多针对不同目的和数据格式的挖掘算法与工具。探究如何应用恰当的挖掘技术，以获取所需度量指标的数据，或可成为未来工作的一个方向，从而推动SPS在实践中的应用。

此外，从SPS研究中综合得出的度量指标和知识并不局限于过程仿真，还可以扩展到一般的软件过程建模。将仿真度量指标作为标准和参考，有助于

进一步推动软件开发人员改进实践中过程数据的收集、治理和应用。

致谢

本研究得到中国国家自然科学基金（No.62072227, No.62202219）、中国国家重点研发计划（No.2019YFE0105500）与挪威研究委员会（No.309494）的联合资助、江苏省重点研发计划（No.BE2021002-2）、江苏省政府间双边创新项目（No.BZ2020017），以及南京大学计算机软件新技术国家重点实验室创新项目和海外开放课题（ZZKT2022A25，KFKT2022A09）的资助。

参考文献

[1] M. I. Kellner、R. J. Madachy 和 D. M. Raffo，“软件过程仿真建模：为何？什么？如何？”《系统与软件杂志》，第46卷，第2期，页91–105，1999年。[2]T. Abdel-Hamid 和 S. E. Madnick，《软件项目动态：一种集成方法》。美国新泽西州上萨德尔河：普伦蒂斯霍尔公司，1991年。[3] R. Ahmed、T. Hall、P. Wernick、S. Robinson 和 M. Shah，“软件过程仿真建模实践调查”，《仿真杂志》，第2卷，第2期，页91–102，2008年。[4] T. Birkh¨olzer、C. Dickmann、J. Vaupel 和 L. Dantas，“基于CMMI框架的交互式软件管理仿真器”，《软件过程改进与实践》，第10卷，第3期，页327–340，2010年。[5] D. Crespo 和 M. Ruiz，“利用多范式仿真建模支持CMMI过程域中的决策”，载于仿真会议，2012年，页1–12。[6] J. Li、M. Li、D. Wu 和 H. Song，“用于可信软件过程管理的综合风险度量与优化模型”，《信息科学》，第191卷，第9期，页47–60，2012年。[7] D. M. Raffo、J. V. Vandeville 和 R. H. Martin，“通过软件过程仿真达到更高的CMM成熟度等级”，《系统与软件杂志》，第46卷，第2–3期，页163–172，1999年。[8] D. Mishra 和 B. Mahanty，“通过外包至低成本目的地，研究软件开发项目的成本、进度和质量”，《企业信息管理杂志》，第29卷，第3期，页454–478，2016。[9] H. Zhang、R. Jeffery、D. Houston、L. Huang 和 L. Zhu，“过程仿真对软件实践的影响：初步报告”，载于*2011*年第*33*届国际软件工程会议（*ICSE*），2011年。[10] H. Zhang、D. Raffo、T. Birkh¨olzer、D. Houston、R. J. Madachy、J. M¨unch 和 S. M. S. Jr.，“软件过程仿真——十字路口上的抉择？”《软件：演化与过程杂志》，第26卷，第10期，页923–928，2014年。[11] O. G´omez、H. Oktaba、M. Piattini 和 F. Garc´ıa，“软件工程测量的系统综述：度量指标的最新进展”，载于*ICSOFT 2006*，第一届国际软件与数据技术会议，塞图巴尔´，葡萄牙，*2006*年*9*月*11*日至*14*日，2006年，第224–231页。[12]C. G. Portobellini、R. D. C. D. Fariapereira 和 J. Luizbecker，“软件工程中的测量：从路线图走向十字路口”，《国际软件工程与知识工程杂志》，第18卷，第01期，第37–64页，2008年。[13] B. Kitchenham，“软件度量怎么了？——一项初步映射研究”，《系统与软件杂志》，第83卷，第1期，第37–51页，2010年。[14]R. Abilio、P. Teles、H. Costa 和 E. Figueiredo，“软件可维护性当代度量指标的系统综述”，载于第六届巴西软件组件研讨会，2012年。[15] E. Kupiainen、M. V. M¨antyl¨a 和 J. Itkonen，“在敏捷和精益软件开发中使用度量指标——工业研究的系统性文献综述”，《信息与软件技术》，第62卷，C期，第143–163页，2015年6月。[16] A. S. Nu˜nez-Varela、H. G. P´erez-Gonzalez、F. E. Mart´ınez-Perez 和 C. Soubervielle-Montalvo，“源代码度量：一项系统性映射研究”，《系统与软件杂志》，第128卷，第164–197页，2017年。[17] A. Meidan、J. A. Garc´ıa-Garc´ıa、I. M. Ramos 和 M. J. Escalona，“软件过程测量：一项系统性映射研究”，《*ACM*计算调查》，第51卷，第3期，第58:1–58:32页，2018年。[18] O. Lindland、G. Sindre 和 A. Solvberg，“理解概念建模中的质量”，《*IEEE*软件》，第11卷，第2期，第42–49页，1994年。[19]B. A. Kitchenham、L. Pickard、S. G. Linkman 和 P. Jones，“评估软件投标模型的框架”，《信息与软件技术》，第47卷，第11期，第747–760页，2005年。[20]B. Kuipers，“定性推理：利用不完整知识进行建模与仿真”，《自动化》，第25卷，第4期，第571–585页，1989年。[21] K. Akingbehin 和 B. Maxim，“软件工程度量的三层模型”，载于第七届*ACIS*软件工程、人工智能、网络及并行/分布式计算国际会议论文集，丛书 SNPD-SAWN ’06。美国华盛顿特区：IEEE计算机学会，2006年，第17–20页。[22] N. E. Fenton，《软件度量：严谨且实用的方法》。International Thomson ComputerPress，2014年。[23] H. Zhang、B. A. Kitchenham 和 D. Pfahl，“对软件过程仿真建模十年发展的反思：一项系统综述”，载于《让全球分布式软件开发取得成功》，国际软件过程会议，*ICSP 2008*，德国莱比锡，*2008*年*5*月*10*日至*11*日，会议论文集，2008年，第345–356页。[24]H. Zhang、B. Kitchenham 和 D. Pfahl，“软件过程仿真建模：事实、趋势与方向”，载于第*15*届亚太软件工程会议*APSEC 2008*，*2008*年*12*月*3*日至*5*日，中国北京，2008年，第59–66页。[25] H. Zhang、B. A. Kitchenham 和 D. Pfahl，“软件过程仿真建模：扩展系统综述”，载于《当今软件过程的新建模概念》，国际软件过程会议，*ICSP 2010*，德国帕德博恩，*2010*年*7*月*8*日至*9*日。会议论文集——

集，2010年，第309–320页。[26] H. Zhang、D. R. Jeffery、D. Houston、L. Huang 和 L. Zhu，“过程仿真对软件实践的影响：初步报告”，载于第*33*届国际软件工程会议论文集，*ICSE 2011*，威基基，檀香山，夏威夷州，美国，*2011*年*5*月*21*日至*28*日，2011年，第1046–1056页。[27] S. Jiang、H. Zhang、C. Gao、D. Shao 和 G. Rong，“面向软件工程教育的过程仿真”，载于*2015*年国际软件与系统过程会议论文集，*ICSSP 2015*，塔林，爱沙尼亚，*2015*年*8*月*24*日至*26*日，2015年，第147–156页。[28] C. Gao、H. Zhang 和 S. Jiang，“构建混合软件过程仿真模型”，载于*2015*年国际软件与系统过程会议论文集，*ICSSP 2015*，塔林，爱沙尼亚，*2015*年*8*月*24*日至*26*日，2015年，第157–166页。[29] H. Gong、H. Zhang、D. Yu 和 B. Liu，“关于验证与确认软件过程仿真模型的系统图谱”，载于*2017*年国际软件与系统过程会议论文集，巴黎，法国，*ICSSP 2017*，*2017*年*7*月*5*日至*7*日，2017年，第50–59页。[30] B. B. Franc¸a 和 G. Travassos，“我们是否已准备好开展软件工程中的仿真研究？”第16卷，第8页，2013年。[31] N. B. Ali、K. Petersen 和 C. Wohlin，“软件过程仿真的工业应用系统性文献综述”，《系统与软件杂志》，第97卷，第65–85页，2014年。[32] D. Pfahl，《过程仿真：软件项目经理的工具？》施普林格·柏林·海德堡，2014年。[33] F. Garc´ıa、M. F. Bertoa、C. Calero、A. Vallecillo、F. Ruiz、M. Piattini 和 M. Genero，“ 构建一致的软件度量术语体系 ”，《信息与软件技术》，第48卷，第8期，第631–644页，2006年。[34] L. Olsina 和 M. de los Angeles Mart´ın，“软件度量和指标本体”，《网络工程杂志》，第2卷，第4期，第262–281页，2003年。[35] S. Robinson，“仿真的概念建模”，载于*2013*年冬季仿真会议论文集：仿真：在复杂世界中做出决策，2013年，第377–388页。[36]——，“仿真的概念建模，第一部分：定义与需求”，《运筹学会杂志》，第59卷，第3期，第278–290页，2008年。[37]I. Standard，“采用 ISO/IEC 15939:2007——系统与软件工程——测量过程”，2009年。[38] B. Kitchenham 和 S. Charters，“软件工程系统性文献综述实施指南，2.3版”，软件工程组、计算机科学与数学学院、基尔大学及杜伦大学计算机科学系，技术报告，2007年。[39] D. Raffo 和 M. I. Kellner，“软件过程仿真建模中的实证分析”，《系统与软件杂志》，第53卷，第1期，第31–41页，2000年。[40] D. Pfahl，“PL-SIM：用于研究汽车行业战略性软件过程改进的通用仿真模型”，载于*W11L*研讨会——第*26*届国际软件工程会议，第2004卷，2004年，第149–158页。[41] H. Zhang、B. Kitchenham 和 D. Jeffery，“组织中采用软件过程仿真的框架”，载于国际软件过程会议，2007年，页320–331。[42] D. S. Cruzes 和 T. Dyb˚a，“软件工程主题综合的推荐步骤”，载于第五届国际实证软件工程与度量研讨会论文集（*ESEM 2011*），加拿大艾伯塔省班夫，*2011*年*9*月*22–23*日，2011年，页275–284。[43]J. M. Corbin 和 A. L. Strauss，“定性研究基础：扎根理论的开发技术与程序”，加利福尼亚州千橡市，*Sage*，*Tashakkori A* 和 *Teddlie C*，第36卷，第100期，页129，1998年。[44] C. Kaner 和 W. P. Bond，“软件工程度量指标：它们衡量什么，我们又如何知道？”*IEEE*计算机学会度量指标，2004年。[45] F. Padberg，“用于评估软件项目调度策略的离散仿真模型”，《软件过程：改进与实践》，第7卷，页127–139，2002年。[46]R. Cherif 和 P. Davidsson，“软件开发过程仿真：基于多智能体的仿真与系统动力学”，载于《基于多智能体的仿真 *X*》：国际研讨会，*MABS 2009*，匈牙利布达佩斯，*2009*年*5*月*11–12*日，修订精选论文，2009年，页73–85。[47] N. Ali 和 K. Petersen，“软件过程仿真的综合流程：现状与行业经验”，2012年9月，页327–336。[48] B. B. Franc¸a 和 N. Ali，基于仿真的研究在软件工程研究中的作用，2020年8月，页263–287。[49] C. M. Abts，“商用现货软件集成成本建模研究”，1997年6月。[50] K. Lum、J. Powell 和 J. Hihn，“飞行与地面系统航天器软件成本估算模型的验证”，2002年。[51]C. Jones，“应用软件度量：确保生产率与质量”，麦格劳-希尔软件工程丛书，1996年。[52] A. A. Frost 和 M. J. Campo，“推进缺陷遏制，实现定量缺陷管理”，2007年。[53]S. Wagner，“缺陷检测技术质量经济学的文献调查”，载于*2006*年 *ACM/IEEE* 国际实证软件工程研讨会论文集，系列：ISESE ’06，2006年。[54] X. Zhou、Y. Jin、H. Zhang、S. Li 和 X. Huang，“软件工程系统性文献综述有效性威胁图谱”，载于第*23*届亚太软件工程会议（*APSEC 2016*），新西兰汉密尔顿，*2016*年*12*月*6–9*日，2016年，页153–160。[55]C. Wohlin，“采用滚雪球法开展第二代系统文献研究”，载于第*20*届国际软件工程评估与评价会议论文集（*EASE 2016*），爱尔兰利默里克，*2016*年*6*月*1–3*日，2016，页15:1–15:6。[56] H. Zhang、M. A. Babar 和 P. Tell，“识别软件工程中的相关研究”，《信息与软件技术》，第53卷，第6期，页625–637，2011年。[57] L. Dong、B. Liu、Z. Li、O. Wu、M. A. Babar 和 B. Xue，“软件过程挖掘映射研究”，载于第*24*届亚太软件工程会议（*APSEC 2017*），2017年。

附录

参考文献

[Aker 17] O. Akerele。“利用系统动力学对敏捷云系统进行仿真建模”。云计算与大数据分析战略工程，页101–117，2017年。[Al E 08] A. Al-Emran、P. Kapur、D. Pfahl 和 G. Ruhe。“仿真最坏情况场景，并分析其在运营发布规划中的综合影响”。页269–281，施普林格，2008年。[Ambr 11] B. G. Ambr´osio、J. L. Braga 和 M. A. Resende-Filho。“软件项目需求活动管理中的决策支持建模与场景仿真”。《软件维护与演化：研究与实践》，第23卷，第1期，页35–50，2011年。[Aran 08] E. Aranha 和 P. Borba。“利用过程仿真评估基于模型的测试方法所缩减的测试设计工作量”。载于软件过程：*2008*年国际会议——让全球分布式软件开发取得成功，页282–293，2008年。[Arau 12] M. A. P. Ara´ujo。“构建支持软件演化计算机模拟研究的模型”。载于*ACM-IEEE*国际实证软件工程与度量研讨会，页281–290，2012年。[Armb 03] O. Armbrust。“利用重复实验获得的实证知识进行软件过程仿真：一个实践示例”。载于国际实证软件工程研讨会，*2003*年。*ISESE 2003*。会议论文集，页18–27，2003年。[Baum 17] T. Baum、F. Kortum、K. Schneider、A. Brack 和 J. Schauder。“利用过程仿真比较提交前评审和提交后评审”。载于*IEEE/ACM*软件与系统过程国际会议，页26–35，2017年。[Berl 03] T. Berling、C. Andersson、M. H¨ost 和 C. Nyberg。“调整仿真模型模板，以便在工业项目中进行测试”。页3–4，2003年。[Cang 08] J. W. Cangussu 和 R. M. Karcich。“利用动态模型评估集成测试和系统测试”。载于国际软件工程与知识工程会议，页436–441，2008年。[Chat 00] B. W. Chatters、M. M. Lehman、J. F. Ramil 和 P. Wernick。“软件演化过程建模：一项长期案例研究”。《软件过程改进与实践》，第5卷，第2–3期，页91–102，2000年。[Chen 06] Y. Chen、G. C. Gannod 和 J. S. Collofello。“软件产品线过程模拟器”。《软件过程改进与实践》，第11卷第4期，页385–409，2006年。[Cher 10] R. Cherif 和 P. Davidsson。“软件开发过程仿真：基于多智能体的仿真与系统动力学”。《计算机科学讲义》，第5683卷，2010年第201期，页73–85，2010年。

[Choi 06] K. Choi、D. Bae，以及 T. Kim。“一种.....的方法一种混合软件过程仿真，使用DEVS形式m”。《软件过程：改进与实践》与实践，第11卷，第4期，第373–383页，2006年。

[Cocc 11] L. Cocco, K. Mannaro, G. Concas，和 M. Marchesi。“模拟看板和 Scrum与瀑布模型的对比以及系统动力学”。载于：*Inter-*全国会议敏捷软件开发大会，第 117–131 页，2011.

[Conc 13] G. Concas、M. M. I. Lunesu、M. Marchesi 和 H. Zhang。“模拟软件维护过程的模拟过程，设有和不设在制品限制”。期刊《软件维护与演进：研究与实践，第25卷，第12期，第1225–1248页，2013年。

[Dan 13] X. H. Dan 和 D. J. Buettner。“用户故事建模完成敏捷软件过程”。载于：国际软件与系统国际会议过程，第 88–97页，2013年。

[Drap 99] A. Drappa 和 J. Ludewig。“定量建模——用于交互式仿真的软件的交互式仿真项目”。期刊《系统与软件杂志》，第 46 卷，第 2-3 期，第 113 页–122，1999年。

[Duga 21] G.-L. Dugarte-Pe˜na, M.-I. S´anchez-Segura, A. de Amescua, F. Medina-Dom´ınguez，和 S. Armenia。“运用系统动力学进行教学关于依赖关系、相关性与系统思维，聚焦于软件过程工作流”。*IET Software*，第 第 15 卷，第 6 期，第 351–364 页，2021年。

[Fate 18] I. Fatema 和 K. Sakib。“使用定性系统动力学开发敏捷团队协作流产力模型”。国际软件进展，《第11卷，第12期，第170–185页，2018.

[Ferr 09] S. Ferreira、J. Collofello、S. Dan 和 G. Mackulak。“理解需求波动性的影响在软件工程中运用分析模型与软件过程仿真化”。期刊《系统与软件》，第82卷，第10期，第1568–1577页，2009年。

[Garo 09] V. Garousi、K. Khosravian 和 D. Pfahl。“一种 customizable 可定制的模式驱动型软件过程仿真模型：设计、校准与应用”。 软件《软件过程改进与实践》，第14卷，第3期，第第165–180页，2009年。

[Garo 16] V. Garousi 和 D. Pfahl。“何时自动化软件测试？一种决策支持方法基于方过程仿真”。 期刊软件：演进与过程，第 28 卷，第 4 期，第 272–285 页，2016.

[Hall 05] T. Hall 和 P. Wernick。“程序切片度量与可演化性：初步研究”。载于：*IEEE* 国际软件可演化性研讨会，第35–40页，2005年。[Hana 02] Hanakawa, Norihiko、Matsumoto、Kenichi、Torii, 和 Koji。“一项基于知识的软件过程仿真模型”。《软件工程年鉴》工程，第 1第4卷，第1-4期，第383–406页，2002年。

[Hana 98] N. Hanakawa, S. Morisaki 和 K. Matsumoto。“一种基于学习曲线的软件开发过程仿真模型opment”。载于：*Forging New Links*,

软件工程会议，*ICSE 98*，*1998*年*4*月*19*日至*25*日。，第350–359页，1998年。[Hous 10] D. Houston 和 M. Lieu。《资源受限的测试与修复周期及测试阶段时长建模》。载于：当今软件过程的新建模概念，国际软件过程会议，*ICSP 2010*，帕德博恩，德国，*2010*年*7*月*8*日至*9*日。会议论文集，第211–221页，2010年。[Hsia 99] P. Hsia、C. Hsu 和 D. C. Kung。《布鲁克斯定律再探：一种系统动力学方法》。载于：第*23*届国际计算机软件与应用会议 *COMPSAC '99*，*1999*年*10*月*27*日至*19*日，美国亚利桑那州菲尼克斯，第370–375页，1999年。[Hurt 15] N. Hurtado、M. Ruiz、E. Orta 和 J. Torres。《利用仿真辅助可用性评估流程管理中的决策》。《信息与软件技术》，第57卷，第509–526页，2015年。[Kahe 01] G. Kahen、M. M. Lehman、J. F. Ramil 和 P. Wernick。《用于政策研究的软件演化过程系统动力学建模：方法与示例》。《系统与软件杂志》，第59卷，第3期，第271–281页，2001年。[Klun 18] J. Kl¨under、F. Kortum、T. Ziehm 和 K. Schnei-der。《帮助团队自助：一项关于冲刺期间相互依赖关系的工业案例研究》。载于：国际以人为中心的软件工程会议，第31–50页，施普林格，2018年。[Kous 07] K. G. Kouskouras 和 A. C. Georgiou。《用于管理软件项目的离散事件仿真模型》。《欧洲运筹学杂志》，第181卷，第1期，第374–389页，2007年。[Lake 03] P. B. Lakey。《用于项目管理的混合软件过程仿真模型》。载于：*2003*年国际软件过程仿真与建模研讨会会议论文集（*ProSim'03*），2003年。[Lehm 10] M. M. Lehman、G. Kahen 和 J. F. Ramil。《长期演化过程的行为建模——若干问题与示例》。《软件维护与演化：研究与实践杂志》，第14卷，第5期，第335–351页，2010年。[Lune 17] I. Lunesu、M. Marchesi、J. M¨unch 和 M. Kuhrmann。《利用测量与仿真理解云中的分布式开发过程》。载于：第*27*届国际软件度量研讨会暨第*12*届国际软件过程与产品度量会议论文集，第1–11页，2017年。[Lune 21] M. I. Lunesu、R. Tonelli、L. Marchesi 和 M. Marchesi。《使用仿真评估敏捷方法中的软件开发风险》。《IEEE Access》，第9卷，第134240–134258页，2021年。

[Mada 00] R. J. Madachy and D. Tarbet. "Case studies in 使用系统动力学进行软件过程建模"。《软件过程：改进与实践》，第5卷，第2-3期，第133–146页，2000年。

[Mada 05] R. Madachy。"业务价值与软件过程的集成建模"。第3840卷，第389–402页，2005年。

[Mada 07] R. Madachy、B. Boehm 和 J. A. Lane。 评估用于 *SISOS* 开发的混合增量过程：研究章节。John Wiley & Sons, Inc.，2007年。

[Mart 00] R. H. Martin 和 D. Raffo。"结合连续模型和离散模型的软件开发过程模型"。《软件过程：改进与实践》，第5卷，第2-3期，第147–157页，2000年。

[Menz 02] T. Menzies、D. Raffo、S. O. Setamanit、Y. Hu 和 S. Tootoonian。《基于模型的 Tru- isms 测试》。载于：*IEEE* 自动化软件工程国际会议，*2002*年。会议论文集。*ASE*，第 183–191 页，2002年。

[Mish 16] D. Mishra 和B. Mahanty。《通过外包到低成本目的地研究软件开发项目成本、进度和质量》。《企业信息管理杂志》，第 29 卷，第 3 期，第 454–478 页，2016。

[Naun 07] P. Naunchan 和 D. Sutivong。《面向基于 COTS 的开发的可调成本估算模型》。载于：软件工程会议，*2007*年。*Aswec 2007*年。澳大利亚，第341–348页，2007年。

[Neu 02a] H. Neu 和 U. Beckerkornstaedt。"通过仿真其底层模型来学习和理解软件过程"。*Kornstaedt*，2002年。

[Neu 02b] H. Neu、T. Hanne、J. M¨unch、S. Nickel 和 A. Wirsen。"用于检查规划的基于仿真的风险降低"。第78–93页，2002年。

[Neu 03] H. Neu、T. Hanne、J. M¨unch、S. Nickel 和 A. Wirsen。"创建用于基于仿真的决策支持的代码检查模型"。发表于：软件过程仿真与建模国际研讨会，2003年。

[Noll 16] J. Noll 和 A. Butterfield。"通过游戏设计教授全球软件开发"。发表于：*2016*年*IEEE*第*11*届全球软件工程国际会议研讨会*(ICGSEW)*，第55–60页，IEEE，2016。

[Oors 18] K. E. van Oorschot、K. Sengupta 和 L. N. Van Wassenhove。"压力之下：迭代周期长度对敏捷软件开发绩效的影响"。《项目管理杂志》，第 49 卷，第 6 期，第 78–102 页，2018年。

[Padb 11] F. Padberg 和 D. Weiss。"面向软件项目的基于模型的调度分析"。载于：计算机软件与应用会议研讨会，第 304–309 页，2011年。

[Park 07] S. H. Park、K. S. Choi、K. Yoon 和 D. H. Bae。"基于SPEM的软件过程模型推导软件过程仿真模型"。载于：*2007*年软件工程会议。*APSEC 2007*。亚太地区，第382–389页，2007年。

[Pfah 00a] D. Pfahl 和 K. Lebsanft。“构建系统动力学仿真模型的知识获取与过程指导：来自软件产业的经验报告”。《国际软件工程与知识工程杂志》，第10卷，第4期，第487–510页，2000年。[Pfah 00b] D. Pfahl 和 K. Lebsanft。“使用仿真分析软件需求波动性对项目绩效的影响”。《信息与软件技术》，第42卷，第14期，第1001–1008页，2000年。[Pfah 01] D. Pfahl、M. Klemm 和 G. Ruhe。“用于软件项目管理教育与培训、集成仿真组件的计算机辅助培训模块”。《系统与软件杂志》，第283–298页，2001年。[Pfah 04] D. Pfahl、M. Stupperich、A. C. Daimlerchrysler 和 T. Krivobokova。“PL-SIM：用于研究汽车行业战略性软件过程改进的通用仿真模型”。《现代非洲研究杂志》，第20卷，第1期，第87–105页，2004年。[Pfah 99] D. Pfahl 和 K. Lebsanft。“将系统动力学建模与描述性过程建模及目标导向度量相结合”。《系统与软件杂志》，第46卷，第2期，1999年。[Raff 00] D. Raffo、W. Harrison 和 J. Vandeville。“协调模型与度量指标以管理软件项目”。《软件过程：改进与实践》，第5卷，第2–3期，第159–168页，2000年。[Raff 02] D. Raffo、W. Harrison 和 J. Vandeville。“软件过程决策支持：运用混合度量、建模与效用框架进行过程模型”。*SEKE 2002*，*意大利伊斯基亚*，*2002年7月15日至19日*，第803–809页，2002年。[Raff 03] D. Raffo 和 S. Setamanit。“运用双向仿真支持软件过程决策”。《国际软件工程与知识工程杂志》，第13卷，第5期，第513–530页，2003年。[Ruiz 01] M. Ruiz、I. Ramos 和 M. Toro。“软件项目动态的简化模型”。《系统与软件杂志》，第59卷，第3期，第299–309页，2001年。[Ruiz 02a] M. Ruiz、I. Ramos 和 M. Toro。“软件过程改进的动态集成框架”。软件质量期刊，第10卷，第2期，第181–194页，2002年。[Ruiz 02b] M. Ruiz、I. Ramos 和 M. Toro。“集成用于基于CMM的软件过程改进的动态模型”。载于：面向产品的软件过程改进国际会议，第63–77页，2002年。[Ruiz 04] M. Ruiz、I. Ramos 和 M. Toro。“运用动态建模与仿真改进商用现货软件过程”。载于：面向产品的软件过程改进，国际会议，*Profes 2004*，日本关西科学城，*2004年4月5日至8日*，会议论文集，第568–581页，2004年。

结合过程仿真与经验数据，以支持软件领域的决策分析开发”。载于：第*14*届国际软件工程与知识工程会议论文集，*SEKE 2002*，意大利伊斯基亚，*2002年7月15日至19日*，第821–833页，2002年。[Rus 03] I. Rus、M. Halling 和 S. Biffl。“支持软件工程中的决策制定与过程仿真和实证研究”。*I*《软件工程与国际期刊》，第1卷，第13期，第531–545页，2003年。[Rus 14] I. Rus、H. New 和 J. M¨unch。“一种系统开发离散事件软件开发过程的仿真模型”。*IEEE*，第1卷，第3期，第1–13页，2014年。[Shuk 12] R. Shukla、M. Shukla、A. K. Misra、T. Marwal 和 W. A. Clarke。“动态软件维护工作量估算建模，采用神经网络、规则引擎和多元回归方 proach”。载于：计算科学与*VIS*应用国际会议*2012* 年。，第 157–16 页

2012年。[Silv 13] L. C. e Silva 和 A. P. C. S. Costa。“决策用于分配信息系统项目人力资源的模型系统项目”。国际期刊项目管理，第31卷，第1期，第 100–10 页

2013年.[Smit 06] N. Smith、A. Capiluppi 和 J. Fern´andez-Rami“基于智能体的开源软件演化仿真过程”。软件过程改进与实践

第11卷第4期，第423–434页，2006年.[Spas 12] B. Spasic 和 B. S.S. Onggo。“基于智能体的仿真：软件开发过程AVL 案例研究”。载于：仿真会议

第1–11页，2012年。[Topi 08] G. Topic、D. Jevtic 和 M. Kunstic。“Petri 网仿真与软件分析开发过程”。收录于：基于知识智能信息与工程系统第*12*届国际会议，*KES 2008*，萨格勒布克罗地亚，*2008年9月3日至5日*，会议论文集，第一部分

第418–425页，2008年。[Tram 16] M. T. I. Trammell、A. Moulton 和 S. E. Manick。“资金波动对软件开发的影响：系统动力学分析”。 工程管理学报，第2卷

第2期，第71–85页，2016。[Turn 06] I. Turnu、M. Melis、A. Cau、A. Setzu、G. Conca 和 K. Mannaro。“建模与仿真采用敏捷实践的开源开发践”。系统架构期刊 *the Euromic*

期刊，第52卷，第11期，第610–618页，2006年。[Uzza13a] M. Uzzafer。“一种应急估算模适用于软件项目”。国际期刊项目管理，第31卷，第7期，第981–99

2013年。[Uzza 13b] M. Uzzafer。“一种用于战略的仿真模型软件项目的管理过程”。*Jou*《系统与软件杂志》，第86卷，第1期，第21页

37，2013年。

[Van 17] K. Van Oorschot、K. Eling 和 F. Langerak：《衡量已知事项以管理未知事项：如何选择新产品开发项目的阶段门时机策略》。《产品创新管理杂志》，2017年。[Wake 04] W. W. Wakeland、R. H. Martin 和 D. Raffo：《运用实验设计、敏感性分析和混合仿真评估软件开发过程的变化：一项案例研究》。《软件过程：改进与实践》，第9卷，第2期，第107–119页，2004年。[Wake 05] W. W. Wakeland、S. Shervais 和 D. Raffo：《将启发式优化作为软件过程仿真模型的验证与确认工具》。《软件过程：改进与实践》，第10卷，第3期，第301–309页，2005年。[Wern 02] P. Wernick 和 T. Hall：《通过结合简单模型仿真全球软件演化过程：一项初步研究》。《软件过程：改进与实践》，第7卷，第3-4期，第113–126页，2002年。[Wern 99] P. Wernick 和 M. M. Lehman：《FEAST/1的软件过程白盒建模》。《系统与软件杂志》，第46卷，第2–3期，第193–201页，1999年。[Xiao 10] J. Xiao、L. J. Osterweil、Q. Wang 和 M. Li：《软件开发过程中的扰动驱动资源重新调度》。载于：《当今软件过程的新建模概念》，国际软件过程会议 *ICSP 2010*，德国帕德博恩，*2010*年*7*月*8*日至*9*日，论文集，第234–247页，2010年。[Zawe 13] A. Zawedde 和 D. Williams：《需求过程改进成功的决定因素》。载于：国际系统动力学会议，2013年。[Zhan 06a] H. Zhang、M. Huo、B. Kitchenham 和 D. R. Jeffery：《软件工程过程定性仿真模型》。载于：第*17*届澳大利亚软件工程会议（*ASWEC 2006*），*2006*年*4*月*18*日至*21*日，澳大利亚悉尼，第391–400页，2006年。[Zhan 06b] H. Zhang 和 B. Kitchenham：《软件工程过程的半定量仿真建模》。载于：软件过程研讨会，第242–253页，2006年。[Zhan 08a] H. Zhang、D. R. Jeffery 和 L. Zhu：《增量开发中测试与修复过程的混合建模》。载于：《让全球分布式软件开发取得成功》，国际软件过程会议 *ICSP 2008*，德国莱比锡，*2008*年*5*月*10*日至*11*日，论文集，第333–344页，2008年。[Zhan 08b] H. Zhang、J. Keung、B. A. Kitchenham 和 D. R. Jeffery：《软件开发过程管理的半定量建模》。载于：第*19*届澳大利亚软件工程会议（*ASWEC 2008*），*2008*年*3*月*25*日至*28*日，澳大利亚珀斯，第66–75页，2008年。[Zhan 09] H. Zhang、B. Kitchenham 和 D. R. Jeffery。“定性与定量软件过程仿真建模：转换与比较”。第345–354页，2009年。[Zhan 12a] H. Zhang、B. Kitchenham 和 R. Jeffery。“迈向可信软件过程模型：一项关于可转换过程建模的探索性研究”。《软件维护与演化：研究与实践》，第24卷，第7期，第741–763页，2012年。[Zhan 12b] H. Zhang、G. Klein、M. Staples、J. Andronick、L. Zhu 和 R. Kolanski。“大规模形式化验证过程的仿真建模”。发表于：国际软件与系统过程会议，第3–12页，2012年。[Zhan 18] X. Zhang、X. Wang 和 Y. Kang。“面向变更的开源软件过程仿真”。*IEEEAccess*，第6卷，第70145–70163页，2018年。[Zhou 12] P. Zhou 和 H. K. N. Leung。“风险管理过程的随机仿真模型”。第1卷，第737–742页，2012年。