# GemPy 1.0: open-source stochastic geological modeling and inversion

Miguel de la Varga, Alexander Schaaf, FlorianWellmann

## 摘要

地下结构的表征涉及广泛的地质调查，从地球流体的储藏研究到地质封存。

基于隐式势场插值方法，构建复杂的全3D地质模型，包括：断层网络、断层-地表交互、不整合断面和dome结构。

Python语言编程，利用Theano库，直接在GPU上执行。

GemPy功能分为：核心库—生成3D地质模型，和附加库—高级科学调查。

以上功能连接到机器学习与Baysian推理框架，因此提供了进行随机地质建模和反演的途径。

## 1引言

地质建模商业软件（如GoCAD, Petrel, GeoModeller），内部工作原理不清楚、拓展方法空间有限、无法适应快速发展的机器学习开源库，如TensorFlow, PyTorch；与其他计算框架整合有限。

Gempy，使用pandas做数据存储和操作；VTK库做3D可视化，2D可视化库Matplotlib；Numpy做数值计算。

## 2 CORE—使用GemPy做地质建模

GemPy的核心功能：从地质输入数据（surface contact points and orientation measurements）构建3D地质模型，以及定义拓扑关系（地层序列和断层网络）。

首先介绍插值算法，然后介绍算法实施及使用GemPy的Python前端执行模型生成和可视化；通过调用一些函数如何构建整体模型（示例代码）。

然后，深入介绍GenPy的框架；如何连接到Theano库，实施高效的向量代数和机器学习，介绍前沿的是随机地质建模和Bayesian反演。

2.1地质建模与势场方法

（1）Potential-field method

Lajaunie et al. (1997)建立的势场方法是GemPy的生成3D地质模拟的核心方法。

不是做时间插值，而是使用一个无量纲参数：标量场数值

使用全局插值代替在感兴趣的底层上分别插值，有2重意义：（1）一个底层位置影响在相同沉积环境下其他底层的位置，相同potential-field的2个底层不可能相交；（2）可以使用底层界面处的数据，打开可用于插值使用到的观测数据的范围。



插值函数作为加权插值（基于universal cokriging）。

（2）调整至地质建模

下面介绍将输入数据代入最终的插值函数**Z**(x0)的步骤。

势场(potential field, Lajaunie et al., 1997)，标量场。

Kriging插值的结果是随机函数。

2.2使用GemPy做地质模型插值

（1）从标量场到地质块模型

结构地质建模的目标是定义地质结构的空间分布，诸如地层、界面和断层。实际上，这些分割通常通过使用体离散(volumetric discretization)或通过描述界面为表面，来完成。

分解区域为离散的3D空间（使用规则网格，如图3）。

目前GemPy仅支持矩形网格，因为计算的标量场是在连续空间，因此允许使用任意类型的网格。这种分割在GemPy中作为lithology block。

第2种分割方法是：定位地层等值面。GemPy使用marching cube算法，类似Gmsh。

（2）整合标量场：沉积序列和断层

大多数地质环境是由被不整合边界分开的沉积相连接而成，受到地震应力引起的位移和变形产生的地层。

Gempy中实施的不连续体包括：不整合与Infinite断层。

不整合模拟是直接的。

断层是通过包含额外的漂移项到Kriging方程组中来模拟的：

2.3GemPy架构

（1）图架构

GemPy架构设计，从基础到适应automatic differentiation (AD)库。

（2）Theano（或TensorFlow）

符号编程语言

机器学习库

## 3 Assets—模型分析和深入使用

3.1可视化

3.2重力正演

3.3拓扑

拓扑，提供描述地质模型中相邻关系的工具，如地层接触或跨断层的连接。

3.4随机地质建模和概率编程

## 4讨论

地质建模的可重复性。

目前，隐式地质建模仅限于一些商业软件，如石油工业的SKUA-GoCAD, Petrel，矿业的GeoModeller, LeapFrog, …商业软件缺乏透明性和不能完全操控一些算法，阻碍着可重复的研究进程。R或Python语言方便科学编程，促进模拟和基于脚本科学的可重复性。GemPy填补地球科学开源生态隐式建模的空白。

隐式建模依赖插值函数，自动化部分或所有构建步骤。近年来，已有很多数学方法解决特殊地质环境的挑战。这些方法的显著优势是：当输入数据修改或增加后，直接重计算整个地质构造。另外，隐式建模还能提供有地质意义的插值函数，即考虑沉积时间或势场来封装不同环境下的地质沉积的要素。

GemPy没有GUI，用户必须提供数据集：坐标和角度数据。因此，对于复杂模型，GemPy可视为后端库，可与其他提供3D GUI的软件联合操作(Gemgis)。

GemPy完全整合于GeoModeller（通过pygeomod库）。

## 参考文献

Lajaunie, C., Courrioux, G., and Manuel, L.: Foliation fields and 3D cartography in geology: Principles of a method based on potential interpolation, Math. Geol., 29, 571-584, 1997.

Miguel de la Varga, Alexander Schaaf, FlorianWellmann. 2019. GemPy 1.0: open-source stochastic geological modeling and inversion. Geosci. Model Dev., 12: 1-32.

# Lajaunie C., Courrioux G., Manuel L.1997. Foliation fields and 3D cartography in geology: Principles of a method based on potential interpolation, Math. Geol. 29: 571-584.

## 引言

描述地质对象的3D几何特征是理解和模拟地质过程的基础。地质模型是由界面(interface, surface)边界约束和由断层(surface)切断的一套formations (volumes)。断层不一定约束formation。

3D制图学的一般目标是：（1）几何正确：是由已知的几何特征拟合的；（2）拓扑上一致：不同地质对象的组成之间的关系得到正确反映；（3）地质上是真实的。

当数据点足够且地质界面相对简单时，经典的地质统计学方法是有效的。而数据稀疏和地质体复杂时，需要为各种地质类型发展特殊的方法。一些插值方法，如Discrete Smooth Interpolation, Bezier surface已应用于几何表面模拟。

Lajaunie, et al. (1997)建立了一种方法：当已知一个或多个界面上的点，以及可获取额外的平面方向(plane orientation)数据。这些方向数据不必属于某个界面，但假设采样来自地质构造(sedimentary plane, foliation, cleavage plane)。地质建模问题就是：构造穿过各界面上已知点的面(surface)，并且与方向数据兼容。

## 适合条件

（1）假设待模拟的面属于一组近似符合叶理构造(foliation)场的平行面。

（2）假设一些方向数据可以转换为矢量数据。因此，从地质学角度看，必须已知某些位置上构造的polarity。

## 基本原理

隐式地质建模的基本原理基于以下思路：空间上定义的标量场，其梯度与方向数据正交。

插值的标量场与一些点有相同的，但未知的标量值（相同界面上的点），此处标量场的梯度或与梯度正交的方向在其他点上是已知的。(The idea is to interpolate the scalar field given that some points have the same but unknown scalar value (points of the same interface), and that either the gradient of the scalar field or a direction orthogonal to this gradient is known at other points.)

最终，模型面由插值后的场的等值面表征。

## 结论

隐式建模的优势是：在一些插值算法中，联合使用不同类型的独立的地质信息，包括：界面上的已知点、其他点上的方向数据以及梯度数据。该方法的另一个特性是：可以模拟一组面，同时考虑不同界面上所有的已知点。

缺点是：模型需要是规则的(regularity)。奇异性(singularity)，比如不可微分的断层（不连续体）需要做特殊的修正；还需要做模型（不确定性）评估。