



波现象与智能反演成像研究组

数据分析中的非线性问题及解决非线性问题的 思想与方法

报告人：王华忠

波现象与智能反演成像研究组 (WPI)

同济大学海洋与地球科学学院，上海

2025年04月08号

目录

- ◆一、概述
- ◆二、数据分析中非线性的本质含义是什么？
- ◆三、地震数据分析中的非线性问题
- ◆四、非线性数据分析的基本思想与方法
- ◆五、结论与讨论



一、概述

- ◆地震数据分析，事实上包括所有的数据分析，基本理论框架都是：把数据视为一个高维的静态数据体，或者随时间变化（也可以随其他标志量变化）的动态数据体。
- ◆从概率统计的角度看，这就是一个高维或一维的随机过程。
 - ◆静态情形下，高维数据体被视为一个空间随机场。
- ◆联合概率密度（函数），理论上，可以描述静态数据体或高维随机过程的统计规律。
- ◆但是，这对于数据分析而言是远远不够的。这也说明了概率统计仅仅提供了统计数据分析的基本理论框架。还要更进一步地发展各种各样的、具体的数据分析方法。

一、概述

◆ 数据分析的根本目的到底是什么？到底想从数据中得到什么？

◆ 狭义地，数据分析的目的是对高维或一维的随机过程进行建模，预测已观测或未观测数据的变化规律。

◆ 建模要建立在数据中包含的线性或非线性结构有一定认识的基础上。因此，数据中蕴含的线性或非线性结构的感知是更基础的。

◆ 广义地，数据分析的目的是Learning From Data。明确地讲，就是从数据中提取出信息，从信息中总结出知识，然后进行最佳的决策或推断。

一、概述

◆数据中到底有什么？

◆狭义地，数据中有模式、有线性或非线性的结构。

◆至于模式、线性或非线性结构的含义是什么，必须由行业内的专业知识来界定。

◆我一直说：只有行业数据分析专家，没有普遍意义下的数据分析专家。

◆必须要成为行业数据分析专家！要对行业数据的底层逻辑要有深刻的认知！

◆否者，学习一堆数据分析基本知识，也不一定能发挥什么作用。尽管学习这些知识是必须的。

◆广义地，数据中一定包含信息。

◆至于信息的内涵是什么，还是要由行业内的专业知识来界定。

一、概述

◆数据中到底有什么？

- ◆可以说，数据中包含的模式、线性或非线性结构，是对数据进行建模并基于模型对未观测或已观测数据进行预测的基础。
- ◆白噪声数据不会包含任何模式、结构，也不包含任何有意义的信息。

◆数据中的模式到底是如何体现出来的？

- ◆它是由随机过程中各随机变量之间的统计相关性决定的。

◆因此，基于上述统计相关性，对数据中蕴含的线性或非线性结构进行感知，并基于它们进行建模，或者说就是对它们进行建模，就构成了数据分析最本质的事情。

一、概述

- ◆ 高斯、线性、平稳理论假设下的数据分析方法，已经非常完善了。
 - ◆ 希望大家真的理解并把握住了高斯、线性、平稳理论假设下的数据分析思想与方法。
 - ◆ 标志性的方法：构建AR模型，进行最佳Wiener滤波。
- ◆ 非平稳、非高斯数据分析方法也已经有很多经典的思想与方法。总结起来：加窗的思想、自适应的思想是应对非平稳数据处理的基本逻辑；Cauchy分布、Laplace分布和Student t-分布等是对非高斯分布的修正。
 - ◆ 标志性的方法是：STFT变换、各种小波变换 --- 非平稳
 - ◆ L1范数或Lp范数下的最佳逼近 --- 非高斯。

一、概述

- ◆但是，非线性问题的解决始终没有提出统一的思想与方法。
 - ◆多项式基函数、径向基函数等线性组合进行的广义线性建模并没有形成非线性数据分析的统一方法。
 - ◆尤其与（强）非平稳（信号）和非高斯（噪声）纠结在一起时，什么是合适的非线性模型，目前完全没有统一的认识。
 - ◆CNN+编码解码这样的黑匣子也很难说是最佳的建模方法选择。
 - ◆非线性问题的最佳建模（逼近）也没有规范性的求解方法。
- ◆总之，非线性、非高斯、非平稳、非因果、非最小相位等情形下的信号分析（或数据分析）才是值得研究的问题。

一、概述

◆值得注意的是：

- ◆我一直认为：复杂情形下信号的建模理论框架，已经涵盖了参数估计的内容，FWI本身没有超出信号分析参数估计理论的内容。
- ◆FWI是假定模型（波动方程）已定，仅仅估计其中包含的参数。当然，它有一些具体问题，但FWI所用的参数估计理论并不特殊。

目录

- ◆一、概述
- ◆二、数据分析中非线性的本质含义是什么？
- ◆三、地震数据分析中的非线性问题
- ◆四、非线性数据分析的基本思想与方法
- ◆五、结论与讨论

二、数据分析中非线性的本质含义是什么？

◆看待非线性问题的视角：

◆数学分析的视角；信号与系统（演化）的视角。



◆数学分析的视角看非线性问题：

◆最本地讲，数学分析中，凡不是线性问题的都是非线性问题。几何上看，直线、平面、超平面表达的问题，就是线性问题。曲线、曲面、超曲面表达的问题就是非线性问题。

◆从数学分析的角度，高维非线性方程（组）代表一个非线性系统。

二、数据分析中非线性的本质含义是什么？

◆看待非线性问题的视角：

- ◆数学分析的视角；信号与系统（演化）的视角。



◆信号与系统（演化）的视角看非线性问题：

- ◆基本地，系统有线性和非线性的。

- ◆任何信号（数据）都是由系统产生的。

- ◆假设噪声为高斯的情形下，线性系统与信号的理论已经非常完善。即便噪声偏离高斯，在噪声广义高斯的情形下，线性系统与信号分析的理论也已经相当地完善。

- ◆随时间演化的非线性系统，抽象地由方程组： $\frac{\partial \mathbf{F}}{\partial t} = \mathbf{F}(\mathbf{x})$ 表示。

二、数据分析中非线性的本质含义是什么？

◆我认为：

◆所谓数据分析中的非线性，就是系统本身的非线性。





二、数据分析中非线性本质含义是什么？

◆ 数据分析中，系统非线性的本质是什么？

◆ 系统非线性的表现形式

◆ 非线性关系：

◆ 输入与输出之间的关系不是直接的比例关系，而是可能涉及平方、立方、指数或其他复杂函数形式。

◆ 多变量相互作用：

◆ 系统的行为可能受到多个变量的共同影响，这些变量之间可能存在交互作用，导致整体效果无法通过单独分析各变量得出。

◆ 反馈机制：

◆ 系统可能存在正反馈或负反馈机制，导致输出能够影响输入，从而形成复杂的动态行为。

◆ 临界点和阈值：

◆ 系统可能在某些条件下突然发生状态变化，表现出非连续的跳跃式行为。

二、数据分析中非线性的本质含义是什么？

◆数据分析中，系统非线性的本质是什么？

◆系统非线性的特性

◆蝴蝶效应：

◆在某些非线性系统中，微小的初始变化可能导致巨大的、不可预测的结果。

◆多稳态性：

◆系统可能存在多个稳定状态，具体表现取决于初始条件或外部干扰。

◆混沌性：

◆某些非线性系统表现出混沌行为，即看似随机的行为，但实际上是由确定性的非线性动力学规律所支配。

◆敏感性：

◆系统对初始条件或参数的变化高度敏感，导致输出的剧烈变化。

二、数据分析中非线性的本质含义是什么？

◆数据分析中，系统非线性的本质是什么？

◆系统非线性在数据分析中的意义

◆复杂模式识别：

- ◆非线性系统通常包含复杂的模式和关系，数据分析需要使用非线性模型（如神经网络、支持向量机等）来捕捉这些模式。

◆预测与控制：

- ◆非线性系统的预测和控制通常比线性系统更具挑战性，需要更复杂的算法和更深入的分析。

◆优化与决策：

- ◆在非线性系统中，优化问题可能涉及多个局部最优解，需要全局搜索算法来找到最优解。

二、数据分析中非线性的本质含义是什么？

◆数据分析中，系统非线性的本质是什么？

- ◆系统非线性的本质在于其复杂性和难以（建模）预测性。
- ◆非线性系统的行为通常是多变量、多反馈、多交互的结果，导致其输出与输入之间的关系远非线性系统的简单和直接。这种非线性在数据分析中既带来了挑战，也提供了揭示系统内在规律和复杂关系的机会。
- ◆到目前为止，我们对非线性系统（演化）的关注不是重点。数据分析中，主要是对非线性结构进行建模，并进行预测。解决各种具体的非线性问题也是重点。粒子滤波方法算是针对非线性系统演化的描述方法。
- ◆无论如何，研究的注意力应转移到非线性系统以及各种非线性问题上来。

目录

- ◆一、概述
- ◆二、数据分析中非线性的本质含义是什么？
- ◆三、地震数据分析中的非线性问题
- ◆四、非线性数据分析的基本思想与方法
- ◆五、结论与讨论

三、地震数据分析中的非线性问题

◆勘探地震中的主要数据类型：

◆叠前地震数据；各种测井数据；岩心测量数据。

◆叠前地震数据：

◆通过波动方程构建起与地下介质弹性参数（变化）之间的因果关系。

◆各种测井数据：

◆通过各种（统计）经验公式构建起与地下介质弹性、电性等参数之间的统计关系。

三、地震数据分析中的非线性问题

- ◆ 勘探地震中的主要数据类型：
 - ◆ 叠前地震数据；各种测井数据；岩心测量数据。
- ◆ 岩心测量数据：
 - ◆ 通过各种（统计）经验公式构建起与地下介质（各种）岩性参数之间的统计关系。
- ◆ 三种主要类型数据的分析中，都存在各自的非线性问题。

三、地震数据分析中的非线性问题

◆叠前地震数据分析中的典型非线性问题：

- ◆基于波场与弹性参数之间的波动方程，进行弹性参数反演，即FWI，是一个典型的（强）非线性问题。
- ◆多次波成像问题，是一个典型的（强）非线性问题。
- ◆从强噪声覆盖下，把弱信号提取出来，是一个典型的（强）非线性问题。
- ◆强噪声+剧变道间时差存在下，对线性和非线性信号建模，尤其是对非线性信号建模，是一个典型的（强）非线性问题。
- ◆噪声较强的情形下，“同因之果”的模式匹配（时差测量），是一个典型的（强）非线性问题。
- ◆还可以继续罗列很多要解决的非线性问题，

三、地震数据分析中的非线性问题

◆ 各种测井数据分析中的典型非线性问题：

◆ 1. 测井曲线之间的非线性关系

- ◆ 测井曲线（如电阻率、密度、声波时、自然电位等）之间的关系通常是非线性的。例如：**电阻率（R）与饱和度（Sw）**：根据阿尔奇公式（Archie 's Equation） $R_t = \frac{R_w}{S_w^n}$ ，电阻率与饱和度的关系是非线性的，且受到孔隙度和水的电阻率 R_w 等因素的影响。其中， n 是饱和度，一般大于1。**密度对孔隙度的非线性响应**：密度测井曲线与孔隙度之间的关系是非线性的，尤其是在多矿物组成的储层中。

◆ 2. 储层非线性特性

- ◆ 储层本身的物理特性通常表现出非线性行为，例如：**渗透率（k）与孔隙度（φ）的非线性关系**：渗透率与孔隙度的关系是非线性的，且受到岩石类型和流体性质的影响。**多相流体的非线性流动**：在储层中，油、气、水三相流体的流动行为是非线性的，且受到饱和度、相对渗透率等因素的影响。**压力与流速的非线性关系**：在非达西流动（如高速流动或非牛顿流体）中，压力与流速的关系是非线性的。

三、地震数据分析中的非线性问题

◆各种测井数据分析中的典型非线性问题：

◆ 3. 测井工具的非线性响应

◆ 测井工具的物理响应通常是非线性的，例如：**电阻率测井的非线性响应**：电阻率测井工具对储层电阻率的响应是非线性的，尤其是在低孔隙度或高电阻率对比的情况下。**声波测井的非线性响应**：声波测井工具对裂缝或复杂储层的响应可能是非线性的，导致测井曲线的解释复杂化。**密度测井的非线性校正**：密度测井工具在高密度或低密度储层中可能会出现非线性偏差，需要进行复杂的校正。

◆ 4. 数据噪声与复杂性

◆ 测井数据在获取过程中可能受到各种噪声和干扰的影响，导致数据分析中出现非线性问题：**测井曲线的非线性噪声**：由于工具误差、地质复杂性或环境干扰，测井数据可能包含非线性噪声，导致传统线性分析方法失效。**多尺度复杂性**：储层的非均质性（如层状异构、裂缝发育）会导致测井数据表现出多尺度的非线性特性。

三、地震数据分析中的非线性问题

◆各种测井数据分析中的典型非线性问题：

◆ 5. 多物理场耦合问题

- ◆ 测井数据分析中，多物理场（如流体力学、电磁场、弹性波等）之间的耦合通常是非线性的，例如：**流体力学与电阻率的耦合**：在多相流体环境中，流体的分布会影响电阻率测井的响应，导致非线性关系。**应力场与声波场的耦合**：在应力敏感储层中，声波测井的响应会受到应力场的非线性影响。

◆ 6. 人工智能与机器学习中的非线性问题

- ◆ 在测井数据分析中，人工智能和机器学习方法（如神经网络、支持向量机等）常被用于处理非线性问题。然而，这些方法本身也面临非线性挑战：**非线性特征提取**：测井数据的非线性特征需要通过非线性模型（如深度学习）来提取。**模型的非线性训练**：神经网络等模型在训练过程中需要处理非线性损失函数和优化问题。

三、地震数据分析中的非线性问题

◆ 各种测井数据分析中的典型非线性问题：

◆ 7. 实际应用中的非线性问题

- ◆ 在实际测井数据分析中，非线性问题还体现在以下几个方面：**Formation Tester (FT) 数据的非线性分析：** FT数据（如压力和流率）通常需要通过非线性模型来分析储层的渗透率和压力系统。**压力传导的非线性效应：** 在复杂储层中，压力传导可能表现出非线性行为，尤其是在多相流体或裂缝储层中。**地震与测井数据的非线性整合：** 将测井数据与地震数据整合时，需要处理两者之间的非线性关系。

◆ 总结

- ◆ 测井数据分析中的非线性问题主要来源于储层的复杂性、测井工具的物理响应以及数据获取过程中的干扰。这些非线性问题需要通过非线性模型、多物理场耦合分析以及高级数据处理技术来解决。

三、地震数据分析中的非线性问题

◆ 岩心测量数据分析中的典型非线性问题：

◆ 孔隙度与渗透率的非线性关系：

- ◆ 孔隙度是岩石中空隙的体积比例，而渗透率是流体通过岩石的能力。两者的关系通常是非线性的，例如卡尔曼方程（Kozeny-Carman equation）描述了渗透率与孔隙度和颗粒大小的非线性关系。

◆ 电阻率与水饱和度的非线性关系：

- ◆ 阿尔奇公式（Archie's Equation）是描述电阻率与水饱和度之间非线性关系的典型模型。公式中，水饱和度的指数通常大于1，导致关系呈现非线性。

◆ 岩石的应力-应变非线性行为：

- ◆ 在岩石力学中，岩石的应力和应变关系在弹性限度内是线性的，但一旦超过弹性限度，就会出现非线性行为，如软化或断裂。

三、地震数据分析中的非线性问题

◆ 岩心测量数据分析中的典型非线性问题：

◆ 多相流体流动的非线性问题：

- ◆ 储层中油、气、水三相流体的流动行为复杂，相对渗透率与饱和度的关系是非线性的，且受到流体性质和流动条件的影响。

◆ 岩心数据的时间效应：

- ◆ 岩石在长时间的加载下可能发生蠕变，表现出时间依赖的非线性行为。

◆ 岩心数据的空间异质性：

- ◆ 储层的性质在空间上可能变化显著，导致岩心数据表现出非线性的分布特征。

◆ 测量误差与噪声的非线性影响：

- ◆ 岩心测量数据可能受到各种噪声和干扰的影响，导致数据分析中出现非线性偏差。

目录

- ◆一、概述
- ◆二、数据分析中非线性的本质含义是什么？
- ◆三、地震数据分析中的非线性问题
- ◆四、非线性数据分析的基本思想与方法
- ◆五、结论与讨论

四、非线性数据分析的基本思想与方法

- ◆数据是由系统产生的。系统是由其所蕴含的模式决定的。模式就是线性或非线性的结构。
- 从随机过程角度看，是随机变量之间的相关性决定了这些线性结构或非线性结构。
- ◆归根结底，非线性数据分析，最根本的问题就是对随机过程进行建模。



- ◆输入、系统与输出之间的关系: $Y_t = H(X_{t-p})$
- ◆系统 $H(*)$ 可以是线性的，也可以是非线性的。
- ◆对系统建模可以抽象为基于后验概率密度函数 $P[Y_t | X_{t-1}, X_{t-1}, \dots, X_{t-p}]$ 的推断问题，无论系统是线性的或是非线性的、平稳的或是非平稳的、估计误差是高斯的或是非高斯的。

四、非线性数据分析的基本思想与方法

- ◆ 数据分析的最抽象表达就应该是求解上述后验条件概率对应的后验期望问题。
- ◆ $H(*)$ 的具体形式也要由后验期望来定。对 $H(*)$ 要有先验认识，先验认识决定 $H(*)$ 的具体形式应该是什么样的。
- ◆ FWI也是求解上述后验期望问题。只是它假设后验期望为Gauss分布的。本质上，Gauss分布假设与系统线性化假设是一致的。
- ◆ 总之，对静态的高维数据体而言，就是对其中结构进行建模。对动态的一维或高维数据体而言，就是对一维或高维数据体中蕴含的模式随时间演化的规律进行建模。建模的目的就是预测、推断、控制。



四、非线性数据分析的基本思想与方法

◆我相信：这就是非线性数据分析的根本理论基础。至于千变万化的具体非线性问题，到底用什么方法来解决，要考验的是自己对基本理论的理解和认识程度。

◆不能变通，就不能创新。

◆下面总结一下有一定代表性的非线性数据分析的基本思想与方法。

四、非线性数据分析的基本思想与方法

◆1、TAR (Threshold Autoregressive) Model

◆可以被描述为:

$$X_t = \sum_{i=1}^k \{b_{i0} + b_{i1}X_{t-1} + \dots + b_{ip_i}X_{t-p_i} + \sigma_i \varepsilon_t\} I(X_{t-d} \in A_i)$$

◆其中, $\varepsilon_t \sim IID(0,1)$; $d, p_1, p_2, \dots, p_k$ 为超参 (正整数)。 $\{A_i\}$ 对 X 的分布范围进行剖分。要求 $A_i \cap A_j = \phi$ 。

◆ $\sigma_i > 0$, 这是什么量? 为什么引入? 分段调节 ε_t 的幅值水平? 统计上, 如何考虑的?

◆ TAR Model 显然是认为 $X_t, t=1, 2, 3, \dots, N$ 可被分成若干段, 每段均可被 AR Model 逼近预测, 每段均为平稳的。

◆但是, 分段的逻辑是什么? 不分的话, 不符合线性假设! 分段的依据 (准则) 是什么呢?

四、非线性数据分析的基本思想与方法

◆2、ARCH (Autoregressive Conditional Heteroscedastic, 自回归条件异方差)

Model

异方差性，方差变化不平稳

◆可以被描述为: $X_t = \sigma_t \varepsilon_t$

◆其中, $\varepsilon_t \sim IID(0,1)$; $\sigma_t^2 = C_0 + b_1 X_{t-1}^2 + \dots + b_p X_{t-p}^2$; $C_0 \geq 0, b_j \geq 0$; 并且期望式 $E(\varepsilon_t X_{t-k}) = 0, k \geq 1, \forall t$ 成立。

◆若把 σ_t^2 取为 $\sigma_t^2 = a_0 + a_1 \sigma_{t-1}^2 + \dots + a_p \sigma_{t-p}^2 + b_1 X_{t-1}^2 + \dots + b_q X_{t-q}^2$; $a_j \geq 0, b_j \geq 0$

◆称 $X_t = \sigma_t \varepsilon_t$ 为广义的GARCH (Generalized Autoregressive Conditional Heteroscedastic) Model。

◆ GARCH模型的基本思想是：当前的方差不仅仅依赖于当前的误差（即冲击），还依赖于过去的方差。这样，模型可以捕捉到方差的延续性和波动性的变化。具体来说，GARCH模型假设当前的条件方差是过去几期方差和当前误差的平方的加权平均。

◆ 统计意义下如何理解ARCH/GARCH Model的含义？逻辑上，如此预测随机过程的模型是如何提出来的？

四、非线性数据分析的基本思想与方法

◆3、非参数化的ARCH Model (Nonparametric Autoregressive Conditional Heteroscedastic (NARCH) Model)

◆可以被描述为：

$$X_t = f(X_{t-1}, \dots, X_{t-p}) + \sigma(X_{t-1}, \dots, X_{t-p})\varepsilon_t$$

◆其中, $\varepsilon_t \sim IID(0,1)$ 。

◆ARCH和NARCH模型是经济领域中常用的模型。但是，其统计意义如何理解？为什么该模型生成的随机序列可以逼近实测的随机序列？

四、非线性数据分析的基本思想与方法

◆ NARCH模型是基于以下考虑提出的：

◆ 克服参数化假设的限制：

- ◆ 传统的GARCH模型依赖于特定的参数化假设，如方差的条件分布通常假设为正态分布或t分布。然而，实际数据中，方差的分布可能更为复杂，无法被这些假设完全捕捉。NARCH模型通过非参数方法，避免了对数据分布的严格假设，使其更具灵活性和适应性。

◆ 捕捉非线性关系：

- ◆ 金融时间序列数据往往表现出非线性特征，如波动聚集、长期依赖等。GARCH模型虽然能够捕捉部分非线性关系，但其线性结构在面对更复杂的非线性模式时可能显得力不从心。NARCH模型通过非参数估计，可以更好地适应和捕捉数据中的非线性异方差性。

◆ 数据驱动的模式构建：

- ◆ 非参数方法通常是数据驱动的，能够自动识别和适应数据中的模式，而无需预先指定模型的具体形式。这种特性使得NARCH模型在处理不同类型的数据时更加灵活和高效，特别是在数据特征不明确或复杂的情况下。

◆ 提高预测准确性：

- ◆ 通过更灵活地捕捉数据中的波动性和非线性关系，NARCH模型有潜力提高对未来波动性的预测准确性。这对于风险管理、资产定价和投资决策等金融应用尤为重要。

◆ 应对高维和复杂数据：

- ◆ 在高维数据或具有复杂结构的数据中，参数化模型可能难以有效捕捉所有相关特征。NARCH模型通过非参数方法，可以更好地处理这些复杂性，提供更全面的波动性建模。

◆ 减少模型误差：

- ◆ 传统GARCH模型可能由于参数化的限制，导致模型误差。NARCH模型通过减少对数据分布的假设，能够在一定程度上降低模型误差，提供更可靠的波动性估计。

◆ 促进方法的多样性：

- ◆ NARCH模型的提出丰富了时间序列分析的方法论，提供了一种新的视角和工具，用于处理条件异方差问题。这种方法的多样性有助于研究人员和从业者根据具体问题选择最适合的模型。

- ◆ 综上所述，NARCH模型的提出主要是为了克服传统GARCH模型在参数化假设、非线性关系捕捉和数据适应性方面的局限性，提供一种更灵活、更具适应性的方法来处理复杂的时间序列数据中的条件异方差问题。

四、非线性数据分析的基本思想与方法

◆4、非参数化的AR Model (Nonparametric Autoregressive (NAR) Model)

◆可以被描述为

$$X_t = f(X_{t-1}, \dots, X_{t-p}) + c \cdot \varepsilon_t$$

◆很显然，这是把非参数化的ARCH Model (NARCH) 中 $\sigma(\cdot)$ 取为 c 时退化产生的。

◆此式似乎可以包含Volterra级数Model。关键是式中的函数 $f(X_{t-1}, \dots, X_{t-p})$ 如何取！

四、非线性数据分析的基本思想与方法

◆5、FAR (Function-coefficient autoregressive) Model

◆可以被描述为：

$$X_t = f_1(X_{t-d})X_1 + \cdots + f_p(X_{t-d})X_{t-p} + \sigma(X_{t-d})\varepsilon_t$$

- ◆显然这是对非参数化的ARCH Model (Nonparametric Autoregressive Conditional Heteroscedastic Model) 中的函数项写出稍微具体一点的形式化得到的。NAR中函数固定成了线性关系式，系数是可变的，与 X_{t-d} 有关。
- ◆提出该式时，统计学上的考虑还要继续澄清。

四、非线性数据分析的基本思想与方法

◆6、加性AR (AAR) Model (Additive autoregressive (AAR) model)

◆可以被描述为:

$$X_t = f_1(X_1) + \cdots + f_p(X_{t-p}) + \varepsilon_t$$

◆这显然是FAR (Function-coefficient autoregressive) Model的再退化。其中已经出现了分段函数叠加的观点，关键是其中的诸函数如何取。

◆如果用样条 (B样条) 来表达诸函数，就是非线性模型的一个很可以实现的选择。值得花时间试一下。这个似乎应该和TAR Model结合，理论逻辑才完整。

◆ARR Model推广应用在高维空间应该更合适。

四、非线性数据分析的基本思想与方法

◆7、常规的AR Model

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} \cdots + a_p X_{t-p} + \varepsilon_t$$

◆这个Model，应该比较清楚了。

◆目前，绝大部分的现代信号分析都是奠基在ARMA/AR/MA模型的基础上。

四、非线性数据分析的基本思想与方法

◆当前存在的问题：

- ◆ 除了常规的AR Model的理论十分清楚外，其它几个模型的背后到底存在什么理论假设不知道，也就不知道它们适用于什么情况。常规的AR Model假设 $X_t, X_{t-1}, \dots, X_{t-p}$ 之间存在线性相关关系，二阶统计量完全刻画了它们的联合概率密度函数，即联合概率密度函数是Gauss的，预测误差 $\varepsilon_t \sim N(0,1)$ 。但是，前述的几个其他的模型为什么能生成可以逼近实际观测序列的随机过程？什么情形下能生成？这个还要从理论上弄明白！
- ◆ 很显然，局部线性化（局部嵌套思想）是对付非线性结构随机序列的基本想法，把常规AR Model推广到分段线性的 $X_t (t=1, 2, \dots, N)$ 的建模预测上，是很直观的想法，据此产生了TAR方法。
- ◆ 用分段样条（B-样条？）函数来实现AAR Model也许是好地建模选择。这个是可以有独创工作的地方。

四、非线性数据分析的基本思想与方法

◆我的困惑：

- ◆到底该用局部数据块的高阶统计量进行建模？还是用局部核函数（径向基函数）进行建模？还是用分段的样条函数进行建模？
- ◆我比较倾向于用后两者进行建模。



◆上述若干模型的参考文献：

- ◆Jiangqing Fan and Qiwei Yao, 2005, Nonlinear Time Series – Nonparametric and Parametric Methods, Springer

目录

- ◆一、概述
- ◆二、数据分析中非线性的本质含义是什么？
- ◆三、地震数据分析中的非线性问题
- ◆四、非线性数据分析的基本思想与方法
- ◆五、结论与讨论

五、结论与讨论

- ◆高斯、线性、平稳情形下的数据分析，已经没有值得研究的空间了。
- ◆要做出点创新的研究工作，必须要针对非高斯、非线性、非平稳情形下的数据分析问题。
- ◆地震波反演成像（FWI）要放在广义信号分析的理论框架下认识，仅仅把它视为是广义信号分析的一个应用场景，尽管在勘探地震中它是一个核心问题。这样才能搭建起与当前大数据、AI需求相衔接的知识架构。

五、结论与讨论

- ◆非高斯、非线性、非平稳情形下的数据分析的核心应该是：
 - ◆用更复杂的模型来捕捉数据中的非线性关系。即寻找数据中复杂的模式和关系，它们不能用简单的线性模型来描述。
 - ◆对静态和动态非线性系统的建模，基于模型进行静态和动态的预测，依然是核心。动态的非线性系统演化结果（混沌、分形、吸引子、平稳态等）的把握不是最核心的事情。
 - ◆与线性系统不同，对非线性系统建模，目前根本没有统一的思想与方法。对于非线性建模问题而言，模型选择、模型确定、模型评价（这三条正是ML/DL的核心问题）都存在很多问题。神经网络是一个典型的黑匣子非线性建模方法，而且不够灵活。

五、结论与讨论



- ◆非高斯、非线性、非平稳情形下的数据分析的核心应该是：
 - ◆所谓的非线性问题建模，可以从不同角度看这个问题，我认为最基本的就是感知到随机过程中**随机变量间的统计相关性（高阶相关性）**，基于这种统计相关性，构建非线性的模型。这基本上是纯数据驱动的建模方法。
 - ◆非线性的回归关系式的建立，是在对数据中存在的相关性有深入理解的基础上，大致确定了非线性的回归关系式的形式，然后进行回归分析。这是一种基于模型先验知识的建模方法。
 - ◆我认为：要先弄清楚**基于统计思想和数学分析方法的非线性问题建模的思想与方法**，然后再去研究神经网络方法。否则，最后的结果是大概率什么也没有学到！

五、结论与讨论

- ◆勘探地震中，非线性建模预测的问题有很多。很多时候，我们并没有刻意地在非线性建模的思想下解决各种具体的非线性问题。
 - ◆譬如：非线性Beamforming；图像配准；非线性时差测量；强噪声弱信号下的去噪；各种层位追踪方法；初至波识别与走时检测方法；速度谱自动解释；高维数据体中非线性信号的建模；地表一致性振幅与相位校正方法等等。还可以列出很多非线性问题。
 - ◆最经典的还是FWI。
- ◆据此，可以看出：针对遇到的具体的非线性问题，需要在理解解决非线性问题根本逻辑的基础上，发展特定的方法与技术。这正是可以体现创新的地方。



**谢谢
欢迎批评指正**