

水下目标辐射噪声的非线性检验*

侯平魁 龚云帆 史习智

(上海交通大学振动、冲击、噪声国家重点实验室 上海 200030)

林良骥

(大连测控技术研究所 大连 116013)

1999 年 9 月 21 日收到

2000 年 4 月 25 日定稿

摘要 由于非线性是混沌产生的必要条件, 所以讨论一个观测时间序列是否具有非线性, 对进一步利用混沌时间序列方法分析观测时间序列很有意义。本文讨论水下目标辐射噪声信号中非线性问题, 运用替代数据方法, 根据实验数据与替代数据之间峭度和自相关两个量的差异性, 分析了水下目标辐射噪声信号中的非线性成分, 得出结论: 水下目标辐射噪声信号中存在非线性成分, 且这种非线性不是由于静态测量函数的非线性引起的, 为利用非线性时间序列分析方法分析水下目标噪声信号提供了理论依据。

PACS 数: 43.30, 43.60

Detecting nonlinearity in the radiated noise of underwater targets

HOU Pingkui GONG Yunfan SHI Xizhi

(State Key Laboratory of Vibration, Shock & Noise, Shanghai JiaoTong Univ. Shanghai 200030)

LIN Liangji

(Dalian Scientific Test and Control Technology Institute Dalian 116013)

Received Sept. 21, 1999

Revised Apr. 25, 2000

Abstract Since non-linearity is a necessary condition for chaotic behavior, the question whether an observed time series consist of nonlinear ingredients is of particular interest for the further analysis of the time series with the chaotic approach. This paper discusses non-linearity in the radiated noise from underwater targets with the surrogate data method. By comparing the differences of a third-order quantity and the third-order auto-correlation between observed data and surrogate data, the non-linearity in the radiated noise is analyzed. The results conclude that there exist non-linearity ingredients in the radiated noise from underwater targets and the non-linearity is not caused by the static nonlinear transformation measurement. The results provide theoretical foundation for the analysis of the radiated noise from underwater targets with the nonlinear time series method.

引言

在现实生活中, 许多自然现象, 如天气的变化、涡流的形成等, 都是貌似随机的, 但它们的实质却都

是由确定的系统所产生的, 由于所包含的非线性的成分, 使之表现得十分复杂^[1]。在发现混沌现象以前, 人们一般以为复杂的信号只会由包含大量自由度的复杂系统产生^[2], 事实上, 当系统为低维混沌

* 国家自然科学基金资助项目 (编号 69772001)

系统时, 由于系统对初始条件的敏感性, 系统会产生貌似随机的复杂信号。然而, 在研究中发现, 如果没有经过检验就假定产生复杂信号的动力系统具有低维的确定性混沌, 直接用 Takens 的延迟坐标嵌入理论对系统进行分析, 包括: 估计系统的自由度以确定嵌入维数、计算李雅普诺夫指数以讨论系统的不确定性等, 这样得出的结论是不可信的。因此, 分析和判断复杂信号是否源于低维混沌对进一步分析信号具有重要的意义。

众所周知, 水声信号是一种复杂且不规则的信号。传统的方法是对采集到的声呐信号运用线性的方法进行处理, 在时域和频域上对信号进行分析, 导出对信号源的描述, 这里认为信号所表现的复杂和不规则性是由于外部随机性作用于动力系统而产生的。非线性时间序列分析理论认为确定性的动力系统内在的非线性会导致系统输出信号的复杂和不规则性^[4]。如上所考虑, 对这样复杂的信号, 其产生机理是什么? 其中是否存在非线性的成分呢? 这就是本文所要讨论的问题。开始, 研究人员是采用对一定时间序列长度内的几何自相似性或以一段时间内序列不确定性的指数增长率来确定混沌, 也有人采用经验模型产生的数据与实验数据进行比较的方法。但是, 在实际应用中, 由于经验模型的系数、阶数确定不易, 所以本文采用替代数据的方法。通过替代数据与实验数据在时间逆下的不对称性及三阶自相关量的比较, 对所采集的水下目标辐射噪声信号进行了分析, 回答了目标信号中非线性存在与否的问题。

1 替代数据方法描述

替代数据 (Surrogate data) 作为检验时间序列中非线性成分的方法是由 Theiler 等人在 1992 年提出的^[5]。该方法的基本思想是首先指定某些特定的线性过程作为零假设 (Null hypothesis), 再实施具体算法由待检验的时间序列出发产生出既满足零假设条件又保留了原序列的诸如均值、方差或者 Fourier 功率谱等性质的替代数据, 然后分别计算待检测序列和替代数据的 Lyapunov 指数、关联维数、相关系数等统计指标, 最后根据原序列和替代数据统计指标的显著性差异水平在一定的置信度内接受或者拒绝假设条件。如果零假设为真, 则说明待检测的序列和替代数据的特征指数无显著性的差异, 因而原来的实验数据与替代数据都是由相同的线性随机过程所产生的; 相反, 零假设不成立, 就意味着待检测序列和替代数据的特征指数之间差异显著, 由于排除了

实验数据是由线性过程决定的可能性, 因而数据当中必定包含有非线性成分在内。

作为检验时间序列非线性因素的重要经验方法, 替代数据方法尽管不能确定引起时间序列非线性的内在机制, 但是当它与某些专门的算法譬如混沌时间序列分析方法相结合使用时^[2], 就可以为检测时间序列非线性的产生机理提供客观依据, 这也是本文采用替代数据方法来检验水声信号中非线性成分的原因。用替代数据方法进行检验的关键之一就是, 产生正确、合适的替代数据序列。下面给出本文采用的两种, 也是较常用的两种零假设和生成相应替代数据的算法。

零假设 1 观测数据是由线性相关的高斯过程所产生。

由传统的信号处理理论可得, 线性相关的高斯过程可以用以下的 ARMA (Autoregressive moving average) 模型表示:

$$x_n = a_0 + \sum_{i=1}^{M_{AR}} a_i x_{n-i} + \sum_{j=0}^{M_{MA}} b_j \eta_{n-j}. \quad (1)$$

为了获得与实验数据相同的均值、方差和自相关函数的替代数据, 一种方法就是从上式出发不断迭代上式, 但是在用实验数据拟合求上式系数的过程中产生的误差会导致最终迭代结果发散, Efron 在文献 6 中对此有详细的描述。Theiler 和 Prichard 采用的对实验数据的 Fourier 变换结果, 进行相位随机化处理的方法非常有效。其处理过程如下:

- (1) 对实验数据进行 Fourier 变换, 得到 $x(k)$ 。
- (2) 相位随机化处理, 得到 $x'(k)$ 。
- (3) 对 $x'(k)$ 求 Fourier 逆变换, 得到 $x'(n)$ 。

$x'(n)$ 就是所要求的替代数据。生成的替代数据保留了原始数据的线性自相关函数, 也就是说, 替代数据与原始数据具有相同的 Fourier 谱, 而非线性自相关性被相位随机化处理去除了。一般把这样生成的替代数据称为随机相位替代数据 (Random Phase surrogate, RP)。

零假设 2 观测数据是由线性相关的高斯噪声经静态非线性变换所产生。

静态非线性变换是指观测或是测量函数具有非线性, 静态或称单调 (static or monotonic) 指的是 t 时刻的观测结果只取决于该时刻动力过程的取值 y^t , 而与以前时刻的值或者导数等无关。

以 $x(n)$ 表示系统的输出时间序列, $S(\cdot)$ 表示一单调静态的测量函数, 则有:

$$S_n = S(x_n). \quad (2)$$

在这种情况下, 尽管产生时间序列 $x(n)$ 的动力学过程是线性相关的, 但是观测到的时间序列 $S(n)$ 却是非线性的。该零假设对应的替代数据将不仅隐含原始数据的线性相关性质, 更为重要的是由测量函数引起原始数据表现出的静态、单调非线性仍然能够为替代数据所反映。比较动力学过程引起的非线性, 观测函数带给时间序列的非线性体现在序列的幅值分布上, 因此, 由符合该零假设的替代数据出发可以检验观测数据中的非线性成分是源于测量函数还是动力学过程本身。

产生该零假设对应的替代数据最常用的方法可以描述如下:

(1) 用伪随机数发生器生成高斯序列 $y(n)$ 。

(2) 以观测数据 $x(n)$ 的秩来重新排列噪声序列, 这样得到的重排噪声序列既遵循观测数据的排列顺序又具有高斯型幅值分布形式。

(3) 对重排噪声序列进行 Fourier 变换和相位随机化处理, 得到新序列 $y'(n)$, 这里和零假设 1 所进行的过程相同。

(4) 按照序列 $y'(n)$ 的秩来排列原始观测数据, 得到所求的替代数据。

从以上过程可以看出, 替代数据与原始序列的幅值分布形式相一致, 因此, 替代数据能够反映出原始序列的静态、单调非线性的性质; 同时, 时间序列 $y(n)$ 和 $y'(n)$ 又具有相同的功率谱, 序列的线性相关性被隐含在产生替代数据的过程当中。按照生成过程称这种替代数据为高斯尺度重排替代数据 (Gaussian-scaled Shuffle surrogate, GS)。在实际的应用中, 以上根据文献 5 中讨论的幅值调整傅里叶方法 (AAFT, the Amplitude Adjusted Fourier Transform method) 给出的替代数据, 由于 $S(\cdot)$ 不一定能够被准确地表述, 会使替代数据是有偏的^[7]。文献 8 中, 作者通过循环调整替代数据的幅值分布和功率谱来达到消除替代数据偏差的目的。

2 零假设的检验

在完成以上步骤, 生成替代数据以后, 我们面临的问题是比较替代数据与实验数据之间的异同, 检验零假设。此时, 怎样来比较二者之间的异同, 选择什么标准来进行比较显得十分重要。

2.1 统计量的选择

统计量选择的正确、合适与否, 直接影响判断结果的优劣, 因此, 选择什么统计量十分关键。对一个混沌系统, 如果具有较强的非线性, 很容易进行检

验。但是, 当面对一个未知系统时, 我们并不知道其特性, 因此, 应选择一个能够检验弱非线性的统计量。在文献 9 中, 作者用时间可逆性来区分线性与非线性过程, 该方法基于线性随机过程的统计量是对称的, 而非线性过程则不然。传统的非线性统计量都是源于两点自相关或信号的功率谱, 它们不包含任何时间方向性的信息, 如: 功率谱。本文根据文献 10 中讨论的高阶统计量和双谱、多谱的应用分析, 选择一个能够较好地反映非线性——时间逆不对称性的三阶统计量, 称它为峭度, 即:

$$\Phi_{rev}(\tau) = \frac{\sum_{n=\tau+1}^N (S_n - S_{n-\tau})^3}{\left[\sum_{n=\tau+1}^N (S_n - S_{n-\tau})^2 \right]^{3/2}}, \quad (3)$$

其中: S_n 代表时间序列, τ 表示时延。

该统计量能够很好地检验时间序列的不对称性。

另外, 作为检验非线性成分的统计量, 三阶自相关也能够达到很好的效果^[7], 即:

$$C_{or3} = \frac{\langle (S_n - \bar{S}) * (S_{n-d} - \bar{S}) * (S_{n-D} - \bar{S}) \rangle}{\langle (S_n - \bar{S})^3 \rangle}, \quad (4)$$

其中: S_n 表示时间序列, d 表示时间延迟, $D = 2d$, \bar{S} 表示均值。

下面用这两个统计量来检验上述零假设 1 和零假设 2。

2.2 检验零假设 1

根据上面所述, 由生成的零假设 1 的替代数据 $x'(n)$, 计算统计量并进行比较。为了使检验的结果具有普适性, 我们在检验中利用 6 类不同水下目标辐射噪声信号的 60 组实验数据产生出 1200 组替代数据, 这样保证产生的数据覆盖所有水下目标的辐射噪声信号, 替代数据与实验数据的比较结果如表 1 所示。表 1 中之所以把不同类目标的计算结果分别列出, 是为了在判断非线性有无的同时, 能够为今后的目标识别工作提供一些依据。事实上, 从结果分析可以看出, F 类目标的三阶自相关量和其它几类有明显的区别, 这为识别工作指明了方向。

从表 1 可以看出, 目标的辐射噪声实验数据与其替代数据在时间逆和三阶自相关两个量之间都存在明显的差异, 对实验数据计算的峭度, 它们都是负值, 而对替代数据求得的峭度, 则是正值。由于替代数据中不存在非线性, 因此根据假设实验数据中一定存在非线性的成分。分别比较各类各组数据的三

阶自相关量, C类数据符号相同,但是量级不同,其他类实验数据与替代数据之间都存在符号的差别,据此判断实验数据中存在非线性成分。因此,我们拒绝零假设 1,也就是说:6类水下目标辐射噪声信号中都存在有非线性。

表 1 零假设 1 检验统计量的比较

		A 类	B 类	C 类
峭度	实验数据	-0.0535	-0.0292	-0.0210
	替代数据	0.0809	0.0310	0.0863
三阶自相关	实验数据	-0.0416	-0.0271	-0.0957
	替代数据	0.8669	0.4829	-0.2261

		D 类	E 类	F 类
峭度	实验数据	-0.0298	-0.0428	-0.0479
	替代数据	0.0236	0.1244	0.0071
三阶自相关	实验数据	-0.0713	-0.0328	0.0016
	替代数据	-5.0108	0.1236	-0.2479

2.3 检验零假设 2

在以上描述中,通过对零假设 1 的检验,我们检测到了水下目标辐射噪声信号中存在非线性,但是这些非线性是不是由测量过程中测量函数的作用而引起的呢?这就是零假设 2 所要回答的问题。零假设 2 认为,观测时间序列中的非线性成分是由观测或测量中的静态非线性变换引起的。

为了能够清楚地表明非线性测量函数的影响,我们给出一阶自回归 (AR) 模型:

$$X_n = 0.7 * X_{n-1} + \eta_n. \tag{5}$$

其产生的时间序列为 X_n 。给定静态非线性函数 S :

$$X3 = S(X) = (X_n)^3. \tag{6}$$

则观测序列为 $X3$ 数据,产生替代数据并根据式 (3) 可以求出和其替代数据的峭度。如表 2 所示:

表 2 测量函数的影响

AR3 替代数据	-2.643	1.213	-2.540	1.373	-1.700
实验数据	0.859	2.119	-0.413	-0.593	-0.835
AR3 替代数据	-2.405	-1.720	3.913	3.787	-2.142
实验数据	-0.854	0.252	1.197	-1.561	-0.962

表中给出 $X3$ 的 20 组替代数据的峭度,其均值为:

$$\langle TimerevAR3 \rangle = -3.655.$$

原始数据的峭度为:

$$Timerev(AR3) = -4.607.$$

由此可以看出,由于静态非线性函数 $S(x) = x^3$ 的作用,使得 AR3 的原始数据与替代数据之间的峭度十分接近。

下面检验水下目标辐射噪声信号中的非线性来源。同假设 1 的检验,仍利用 6 类目标、60 组实验数据产生 1200 组替代数据来分析辐射噪声信号。计算结果如表 3 所示:

表 3 零假设 2 检验统计量的比较

		A 类	B 类	C 类
峭度	实验数据	-0.0237	-0.0253	-0.0162
	替代数据	0.1998	0.0619	0.0856
三阶自相关	实验数据	-0.0608	-0.0377	-0.0921
	替代数据	0.0019	0.0003	-0.0040

		D 类	E 类	F 类
峭度	实验数据	-0.0364	-0.0379	-0.0395
	替代数据	0.1210	0.0188	0.0777
三阶自相关	实验数据	-0.0636	-0.0394	-0.0025
	替代数据	0.0017	0.0005	0.0034

从表 3 可以看出,比较峭度、三阶自相关两个量,实验数据与替代数据之间存在显著的差异。对实验数据,计算的峭度均为正值,而替代数据求得的峭度为负值,存在差异。比较实验数据与替代数据的三阶自相关量可得:除 C 类外,其他各类都存在符号上的差异,而 C 类实验数据与替代数据的三阶自相关量在量级上存在差异。根据替代数据理论我们拒绝零假设 2,也即:水下目标辐射噪声信号中存在的非线性成分,不是由静态非线性的测量函数所引起的,而是系统本身非线性的表现。这对我们进一步利用非线性时间序列方法分析水下目标辐射噪声信号,有重要意义。

3 结论及展望

从以上讨论可以看出,替代数据方法的优点在于能够对所提出的假设在一定的置信度范围内作出接受或拒绝的判断,从而为进一步的信号特征分析提供方向性的指导。本文通过对两个零假设的替代数

据与实验数据的比较, 得出结论: 水下目标噪声信号中确实存在非线性特性, 并且非线性受到测量函数以外因素的影响, 即信号源是一个非线性的动力系统。通过实验数据的比较, 为进一步利用非线性时间序列分析方法对水下目标噪声信号进行处理和识别提供了理论依据。非线性成分的检验只是第一步, 重要的是利用水下目标辐射噪声信号的非线性来识别水下目标, 这也是作者的下一步工作。

尽管替代数据方法检验时间序列的非线性是一种新方法, 在很多领域(如: ECG、EEG、股票市场及生物系统)都得到了很好的实验效果, 但是, 它仍存在许多不足之处。首先替代数据是由假设出发, 在实验数据的基础上产生的, 因此验证结果的好坏受假设的影响比较大。其次, 目前所提供的假设覆盖面不够广, 有些信号的性质还不能通过假设替代数据来进行检验(如: 信号的平稳性检验)。

最近, 在非线性领域里, 许多学者都在致力于拓宽能够利用替代数据检验的零假设的覆盖面, 随着零假设的增多, 信号的非线性性质会更多地被检验、验证。

参 考 文 献

- 1 Haykin S & Principe J. Making sense of a complex world. *IEEE Signal Processing Magazine*, 1998; 5: 66—81
- 2 龚云帆. 非线性动力系统混沌时间序列分析. 西安: 西安交通大学博士论文, 1998
- 3 Abarbanel H D I. Analysis of observed chaotic data. New York: Springer, 1996
- 4 Kantz H, Schreiber T. Nonlinear time series analysis. Cambridge: Cambridge University Press, 1997
- 5 Theiler J, Eubank S, Longtin A *et al.* Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, 1992; 58: 77—83
- 6 Efron B. The jackknife, the bootstrap and other resampling plans. *SIAM, Philadelphia, PA*, 1982
- 7 Schreiber T. Interdisciplinary application of nonlinear time series methods. *Physics Reports*, 1999; 308: 1—64
- 8 Schreiber T, Schmitz A. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.*, 1996; 77(4): 635—638
- 9 Diks C, van Houwelingen J C, Takens F. Reversibility as a criterion for discriminating time series. *Phys. Lett. A*, 1995; 201: 221—233
- 10 Subba Rao T, Gabr M M. An introduction to bispectral analysis and bilinear time series models. *Lecture Notes in Statistics Vol. 24*, New York: Springer, 1984