한국추진공학회 2011년도 추계학술대회 논문집 pp.241~245 2011 KSPE Fall Conference

엔진 시험 데이터에 대한 시계열 분석

김일두*† · 윤현걸* · 임진식*

Time Series Analysis of Engine Test Data

Ildoo Kim*† · Hyun-Gull Yoon* · JinShik Lim*

ABSTRACT

In an engine test, data are collected in a form of a time series. Usually only the time average of a time series is interesting to engineers while its stochastic fluctuation is being ignored. In this paper, we collect pressure and fuel flux data from an air-breathing engine test and analyze their fluctuations using the multiscale sample entropy analysis, which is suggested as a measure of the complexity of a time series. It is shown that different physical quantities indeed have different complexities at each timescales, suggesting a possibility of an instantaneous tool which evaluates the engine test.

초 록

엔진 시험과정에서 데이터는 시계열 형태로 수집된다. 보통 그러한 시계열들의 섭동보다는 시간 평균에 더 관심을 가진다. 본 논문에서는 공기 흡입식 엔진의 시험에서 측정된 압력과 유량 데이터의 섭동에 시계열의 복잡성의 척도로 제안된 개념인 multiscale sample entropy라는 분석법을 적용해본다. 분석 결과, 서로 다른 물리량은 각각의 시간척도에서 다른 복잡성을 가진다는 것을 보였고, 이를 잘 이용하면 엔진 시험의 성패 여부를 즉각적으로 알려주는 도구를 만들 수 있을 것이다.

Key Words: Engine Test (엔진 시험), Time Series Analysis (시계열 분석), Sample Entropy, Air-breathing Engine (공기 흡입식 엔진)

1. 서 론

시계열(time series)은 특정한 값의 시간에 따른 변화를 기록한 것을 통칭하며, 모든 종류의 동역학에서 사용된다. 일기예보나 화학 양론 등 자연과학에서부터 의과학(심장박동)이나 경제학(주가) 등 시계열은 데이터 정리 방법으로 두루 쓰이고 있다.

새로운 추진기관을 개발 할 때 필수적인 추진기관의 시험평가 데이터 역시 주로 시계열로 수집된다. 이 경우 주로 압력, 온도 등의 물리량의 시계열이며, 측정에는 일반적으로 필요한 정보(일정 시간 동안의 평균 압력 등) 이외에도 여러 이유에 의한 섭동(fluctuation)이 포함된다. 이러

* 국방과학연구소 제1기술연구본부 5부
† 교신저자, E-mail: ildoo.kim@add.re.kr

한 시계열을 확률시계열(stochastic time series) 라 한다. 섭동이 기억력 없는 Markov chain인 경우와 같이 (분자 단위의 브라운 운동 등) 규칙성이 전혀 없는 경우도 있다. 반면 유체에서와 같이 섭동이 지배방정식인 Navier-Stokes 방정식의 비선형성에서 나오는 경우, 무작위해 보이지만 일정한 통계 기법을 적용했을 때 규칙이 발견되는 경우도 있다. 대표적인 예가 난류로서, 난류에서 측정되는 유체의 속도는 불규칙하고 예측이 불가능해 보이지만 측정되는 섭동의 크기와 빈도는 푸리에(Fourier) 공간에서 멱급수 분포를 따르며, 그 지수는 지배방정식을 풀어 유도할 수 있다.

고전적으로는 푸리에 변환과 n차 모멘트를 주로 이용하던 시계열을 분석하던 것이 1950년대 이후 비선형 물리학과 카오스 이론이 발전하면서 fractal dimension이나 Lyapunov exponent, Kolmogorov-Sinai entropy 등의 시계열의 복잡성의 척도가 제시되어왔다[1]. Approximate Entropy (ApEn) 역시 시계열의 복잡성을 측정하기 위해 제안된 여러 척도들의 하나로서, 1991년에 미국의 S. Pincus에 의해 제안되었다[2]. 본 논문은 엔진 시험을 통해 얻어진 데이터에 ApEn와 유사한 sample entropy (SampEn)를 계산해 봄으로서 어떤 새로운 특성이 발견될 수 있나 알아보려고 한다. 2장에서 실험 데이터에 적용해볼 SampEn를 비롯한 계산 방법에 대해 서술하고 3장에서는 실제 데이터에 적용된 결과를 제시할 것이다.

2. 시계열 분석 방법

엔진 시험을 통해 얻어지는 데이터는 열, 온도 등 여러 가지가 있으나 본 논문에서 사용하는 분석방법은 특정 물리량이나 단위에 국한되지 않은 일반적인 비차원 수열에 적용될 수 있는 방법이다. 따라서 계산을 하기 전에 시계열을 먼저 정규화 시켜서 단위를 제거하는 과정이 필요하다. 시험을 통해 우리가 $\{v_0, v_1, \ldots, v_{N-1}\}$라는 시계열을 얻었다고 하자. 이 시계열은 N개의 데이터 포인트로 구성되어 있으며 각각의 데이터 포인트는 sampling rate t_0의 시간 간격을 두고 측정되었다. 우리는 이 시계열의 평균 $\langle v \rangle$와 표준편차 $\sigma[v]$를 다음과 같이 정의할 수 있다.

$$< v > = \frac{1}{N}\sum_{n=0}^{N-1} v_n,$$

$$\sigma[v] = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}(v_n - < v >)^2}.$$

각 시계열의 각 요소 v_n을 $(v_n - \langle v \rangle)/\sigma[v]$로 대체함으로서 정규화 할 수 있다. 이후로 언급되는 모든 시계열은 정규화된 것으로 한다.

이렇게 얻어진 정규화 된 시계열에 대해 먼저 자기상관함수를 계산한다. 정규화 된 시계열에 대해 자기상관함수 AutoCR을 다음과 같이 간단히 계산할 수 있다.

$$\mathrm{AutoCR}(s) = \frac{1}{N-s}\sum_{n=0}^{N-s-1} v_n v_{n+s}$$

자기상관함수는 일반적으로 계의 기억효과를 추정하는 데 쓰인다. 자기상관함수가 s가 증가함에 따라 빠르게 0으로 수렴할수록 시계열은 완전한 무작위함수와 같다. 유체역학에서는 섭동이 완전한 무작위가 아니라 거대 구조(large-scale structure)의 영향을 받기 때문에 eddy turnover time이 자가상관함수를 통해 나타나기도 한다.

자가상관함수 다음에는 SampEn를 계산한다. 앞서 말한 ApEn는 예측가능성을 통해 시계열의 복잡성을 나타내는 척도이다. ApEn는 엄밀하게는 길이가 무한대인 시계열에 대해 정의되어 있고 시계열의 길이가 바뀌면 값이 달라지는 단점이 있다. 이러한 단점을 극복하기 위해 길이에 대한 의존성을 줄인 SampEn가 제안되었다[3]. 두 개념은 길이의존성 이외에는 정확히 같으며, 이들의 정의와 의미에 대해서는 많은 논문들이 나와 있다 [2, 3, 4, 5]. SampEn의 유용성에 대해서는 심장박동 시계열에서 환자가 가진 장애의 종류를 판별한 논문이 있다[6]. 본문에서는 개략적으로만 설명하도록 한다. SampEn은 시계열의

예측가능성을 측정하여 복잡성을 추정하기 위해 고안된 개념이다. 이는 Bayes의 법칙

$$P(B|A)=P(A|B)\frac{P(B)}{P(A)}$$

으로부터 시작하여 A를 {시계열 내에서 길이가 m인 시계열의 일부분이 *같은* 값을 가지며 반복되는 경우}, B를 {그 반복되는 시계열이 *같은* 데이터 포인트로 이어지는 경우}로 정의한다. 데이터가 같다는 의미는 비교하는 두 값의 차가 미리 정한 허용범위 r보다 작을 경우를 의미한다. 엄밀한 표현으로 A는 시계열 $\{v_0 \ldots v_N\}$ 안에서 서로 다른 n과 l에 대하여 부분집합 $\{v_n \ldots v_{n+m-1}\}$과 $\{v_{n+l} \ldots v_{n+l+m-1}\}$이 존재하며 이 때 모든 $k(<m)$에 대하여 $|v_{n+k}-v_{n+l+k}|<r$이 성립될 때가 되며 B는 A가 일어났을 때 추가적으로 $|v_{n+m}-v_{n+m+l}|<r$ 이 성립할 때가 된다. 의미적으로 A는 {시계열 내에서 한번 일어난 패턴이 한 번 더 반복될 경우}이고, B는 {A가 일어났을 때 이후 이전과 같은 결과가 따라오는 경우}가 된다. 진리집합 B는 진리집합 A에 포함되기 때문에 $P(A|B)=1$이며 조건부 확률은

$$P(B|A)=\frac{P(B)}{P(A)}=\frac{N_B}{N_A}$$

이 된다. 여기서 N_B와 N_A는 각각 B와 A의 경우의 수가 된다. SampEn는 이 조건부 확률의 음의 로그로 정의된다.

$$\mathrm{SampEn}(r,m;N)=-\ln(N_B/N_A)$$

SampEn는 A가 일어날 경우 B가 일어날 확률이 클수록 (시계열이 예측이 가능할 경우) 0에 가까운 양의 값을 가지게 되고, 시계열이 전혀 예측가능하지 않을 경우 무한대의 값을 가지게 된다. SampEn는 ApEn에서 나타나는 N 의존성을 줄이기 위해 만들어졌기 때문에 이론적으로는 N과는 무관하나 N이 감소할수록 $\sqrt{N}$에 비례하여 분산이 증가한다.

마지막으로 시계열을 일정 간격으로 겹치지 않게 평균을 내어 거칠어진(coarse-grained) 시계열

$$v_i^{(\tau)}=\frac{1}{\tau}\sum_{j=0}^{\tau-1}v_{i\tau+j}$$

을 만들 수 있다. 여기서 τ는 coarse graining time이다. 일종의 low-pass filter로 운용되는 과정을 통하여 하나의 시계열에 따라 다른 시간척도에 해당하는 새로운 시계열을 만들 수 있고 이에 대해서도 SampEn를 계산할 수 있다 (multiscale sample entropy analysis, MSE분석법).

3. 계산 결과

본 논문에서는 공기 흡입식 엔진 시험에서 수집된 시계열을 분석하였다. 엔진 시험에서 수집된 연소실 압력[kgf/cm^2] P와 연료 유량[kg/s] F를 1 kHz로 측정한 시계열을 4회에 걸쳐 수집하였고 이에 대하여 MSE 분석을 수행하였다. 통상적으로 사용되는 대로 m=2와 r=0.15σ를 사용하였다[5].

Figure 1에서는 첫 번째 시험에서 측정된 시계열의 자기상관함수를 볼 수 있다. 검은 색 실선으로 나타난 무작위 함수의 자기상관함수는 s=0에서 델타함수의 꼴로 나타나는 반면 P와 F의 경우는 자기상관이 진동하면서 감소하여 s=25 근처에서는 거의 0에 가까워진다. 특징적으로 자기상관함수가 0을 기준으로 진동하는데 이는 1 kHz보다 빠른 물리현상이 있음을 암시한다. 또한 25 ms 정도인 자기상관 감소시간을 통해 연소실 유동의 거대구조의 eddy turnover time을 유추할 수 있다.

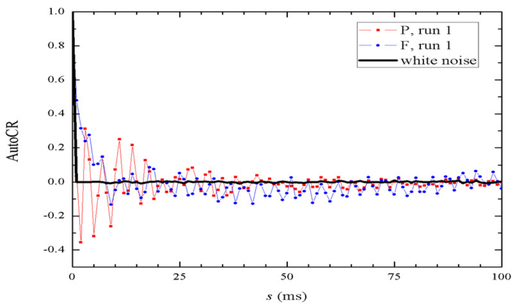

Fig. 1 Autocorrelation functions

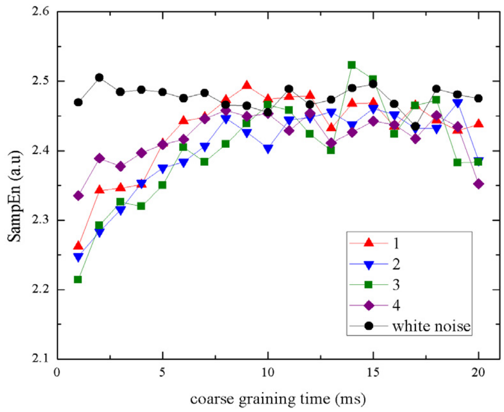

Fig. 2 Multiscale sample entropy of P

Figure 2에서는 측정된 연소실 압력 P에 대해 MSE 분석을 수행한 결과이다. 검은색 원은 인위적으로 만들어진 무작위 함수에 대해 같은 분석을 수행한 결과이다.

4번에 걸친 실험 모두에서 P는 무작위 함수보다 예측가능성이 높은 것으로 계산되었다. 하지만 10 ms정도의 시간척도나 그 이상에서는 무작위함수과 거의 비슷한 값이 나왔는데 이는 시계열의 질서가 대부분 짧은 시간척도에 몰려있다고 해석할 수 있다. 이는 앞서 계산한 자기상관함수와도 일치하는 결과로 1 kHz 이상으로 압력을 측정했을 때 물리적으로 의미 있는 현상을 보일 수도 있다는 것을 암시한다.

마지막으로 Fig. 3에서는 F의 MSE 분석을 수행한 결과이다. 4번의 실험 모두 무작위 함수보다는 모든 시간

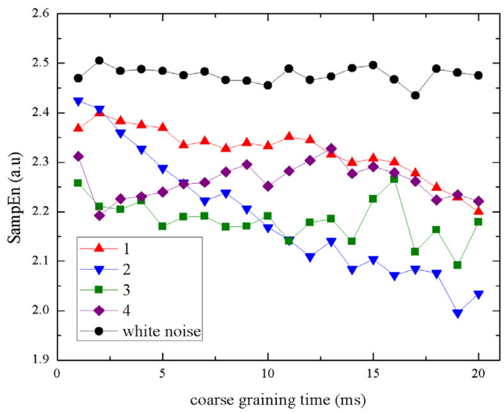

Fig. 3 Multiscale sample entropy of F

척도에서 작은 값을 보인다. 4회에 걸쳐 일관성 있는 분석 결과가 나온 P와는 달리 F에 대한 시계열 분석 결과는 시계열에 대한 분석 결과는 회차에 따라 다소 차이를 보인다. 1, 2번 실험의 경우 τ가 증가함에 따라 SampEn 값은 감소하는 경향이 나타난다. 이는 작은 τ에서 질서도가 크게 나타난 P와는 상반된 결과로서, 빠른 섭동보다는 느린 섭동이 더 물리적으로 중요함을 암시한다. 반면 4번 실험의 경우는 τ의 변화에 따른 SampEn 값의 변화가 거의 없는 "척도 없는 (scale-free)" 특징을 보인다고 할 수 있다. 왜 같은 물리량에 대한 반복 실험에서 다른 결과가 나왔는지에 대하여서는 더욱 연구해 볼 필요가 있다.

4. 결 론

엔진 시험에서 얻어진 시계열에 대해 자기상관함수와 SampEn를 계산해보았다. 엔진 데이터는 1 ms 간격으로 측정이 되어 수집되었으며, 얻어진 시계열을 분석한 결과 최소 시간 단위인 1 ms 보다도 짧은 시간척도를 가진 빠른 섭동이 있을 가능성이 제기되었다. 이러한 추정은 엔진의 구조 해석 등에도 유용하게 참고 될 수 있을 것이다.

또한 차원이 다른 두 물리량에 같은 분석법을 적용했을 때 확연히 다른 결과가 나옴을 보였다. 향후 더 연구할 경우, 수집되는 시계열 데이터만으로 실시간으로 얼마나 엔진 연소가 잘 되고 있는지를 판단할 수 있는 도구를 만들 수 있을 것이다.

참 고 문 헌

1. Hilborn, Robert C., "Chaos and Nonlinear Dynamics", 2/e, Oxford University Press, USA, 2001
2. Pincus, Steven M., "Approximate entropy as a measure of system complexity", Proc. Natl. Acad. Sci. USA, 88, 2297, 1991
3. Richman, Joshua S. and Moorman J.

Randall, "Physiological time-series analysis using approximate entropy and sample entropy", Am. J. Physiol: Heart Circ. Physiol., 279, H2039, 2000
4. Kim, Ildoo, "Experimental Studies on 2D fluid", Ph.D. Dissertation, University of Pittsburgh, 2011
5. Costa, M, Goldberger, Ary L., Peng, C.-K., "Multiscale entropy analysis of biological signals", Phys. Rev. E, 71, 021906, 2005
6. Costa, M, Goldberger, Ary L., Peng, C.-K., "Multiscale Entropy Analysis of Complex Physiological Time Seires", Phys. Rev. Lett., 89, 068102, 2002