복합재 허용치 생성을 위한 점추정 통계분석기법에 관한 연구

Study on the Single-Point Statistical Analysis Techniques to Generate the Composite Material Allowables

이승윤[1]*
한국항공우주연구원[1]

초 록

원재료 성분, 프리프레그 제작공정, 재료 취급, 부품 제작 기술, 적층순서, 환경조건, 그리고 시험방법 등 여러 요인으로 인해 복합재 물성치는 일반적으로 금속재에 비해 변동성이 높다. 따라서, 이러한 높은 변동성을 고려하기 위해서는 엄격한 통계분석기법을 적용하여 복합재의 기계적 물성에 대한 설계 허용치를 계산해야 한다. 1990년대 후반 미국에서는 FAA와 NASA를 중심으로 표준화된 절차에 따라 항공기 설계에 적용가능한 복합재 물성 데이터베이스를 구축하고 공유하기 위한 프로세스를 개발하기 시작하였고, 현재 NCAMP를 중심으로 복합재 데이터베이스 구축 작업이 진행되고 있다. NCAMP는 기본적으로 CMH-17에서 채택한 통계분석기법을 이용하여 허용치를 계산하고 있으며, 본 논문에서는 복합재 허용치 계산 기법 중 점추정 방법을 이용한 통계분석 기법과 그 적용에 대해 논하고자 한다.

1. 서 론

섬유강화 복합재료는 금속재료에 비해 본질적으로 높은 변동성(variability)을 갖는다. 시험방법과 관련된 변동성은 수년 간의 연구와 표준화 작업을 거쳐 최소화되고 있다. 그럼에도 불구하고 복합재가 이방성으로 인해 금속재에 비해 복잡한 기계적 거동을 나타내고 또한 시험 비용이 상대적으로 높기 때문에, 현재까지 구축된 데이터베이스가 금속재에 비해 적은 것이 현실이다. 따라서, 복합재의 각 기계적 물성에 대해 합당한 설계 허용치를 부여하기 위해서는 이러한 변동성을 고려하여 정확한 통계분석기법을 적용해야 한다.

본 논문에서는 재료 허용치 생성을 위한 마지막 단계인 통계분석기법 중 점추정 방법을 중심으로 그 종류와 절차, 그리고 허용치 계산에서 유의할 점에 대해 논하고자 한다. 특히 주의할 점은, 분석 과정에서 공학적 판단 또는 경험에 근거하여 통계적 진단 결과를 종합적으로 평가해야 한다는 것이다. 이러한 복합재 재료인증에 사용되는 통계분석기법들은 CMH-17[1]가 수락한 방법들이다.

2. 통계분석 기법 개요

통계분석으로 재료 허용치를 구하기 전에 반드시 만족해야 하는 주요 전제 조건 중 하나는 적절하게 설계된 표준화된 시험 프로그램을 통해 요구되는 물성을 획득하는 것이다. 이러한 조건이 만족되지 않으면 어떠한 통계분석 기법도 의미있는 결과를 도출하기에 충분하지 않다. 특히 강도 물성 시험에 있어 시편별 파단모드의 일관성과 적절성을 평가하는 것이 중요하다.

복합재의 기저값(basis value)을 계산하는 통계분석 기법은 획득된 데이터의 분포 특성에 따라 합동 방법(pooling method)과 점추정 방법(single-point method)으로 나누어진다. 본 논문에서는 점추정 방법에 대해 중점적으로 다룬다. 기저값 생성을 위한 점추정 방법에 대한 순서도가 Fig. 1에 주어져 있다.

환경조건의 개수가 2개 이상이 되지 못하거나, 환경조건 간 파단모드의 일관성이 없거나, 정규성 조건을 만족하지 못하는 경우 등 합동 요건이 충족되지 못하면 점추정 방법을 이용해야 한다. 점추정 방법에서는 주어진 데이터가 재료 뱃치와

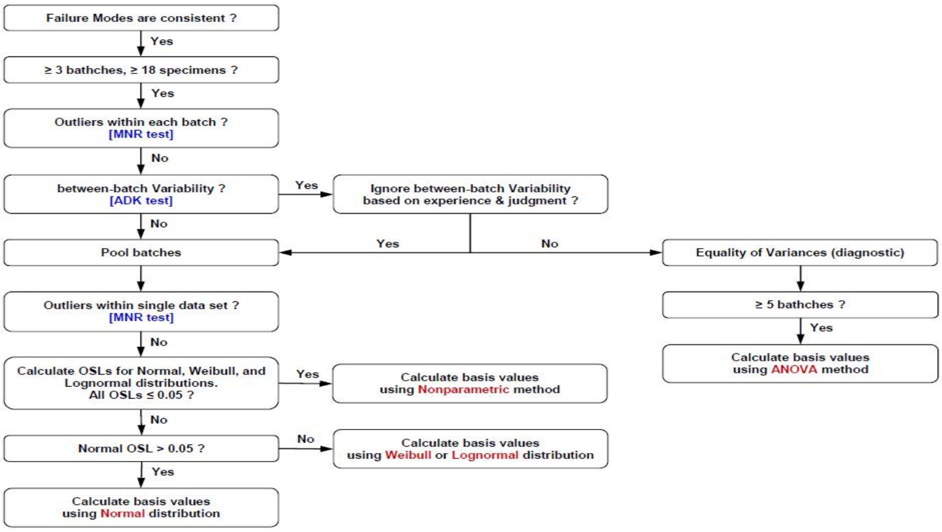

Fig 1. Flowchart of Single-Point Method

같은 자연스러운 방식으로 분류되는지 여부에 따라 다른 접근법이 사용된다. 분류될 수 없거나 혹은 분류된 데이터 집단 간의 차이가 무시할만한 데이터를 비구조적 데이터라고 하며, 그렇지 않은 데이터는 구조적 데이터라고 한다. 비구조적 데이터는 정규분포, Weibull 분포, 또는 로그정규분포를 이용하여 분석한다. 이 중 적절한 모델이 없다면, 비모수적 방법으로 기저값을 결정한다. 구조적 데이터는 회귀분석 및 분산분석(ANOVA) 등을 이용하여 분석한다.

3. 점추정 방법

3.1 정규화 및 이상치 식별

각 환경조건에 대해 각각 분석을 시작하기 전에, 모든 섬유-의존성 물성 데이터는 각 시편의 섬유체적율에 대해 정규화(normalization) [2]를 수행한다.

각 뱃치 내의 이상치를 MNR(maximum normed residual) 검정 및 시각적인 조사를 통해 식별한다. 이상치가 발견되면 공학적인 판단이나 경험에 따라 처리한다. 이상치 데이터 값이 제거되면, 남아있는 데이터의 양이 뱃치와 시편의 최소 개수 요건을 만족해야 한다.

3.2 뱃치 간 변동성

뱃치 데이터 간에 변동성이 무시할만큼 작은지 조사하기 위해 ADK(k-sample Anderson Darling) 검정을 수행한다. ADK 검정은 두 개 이상의 데이터 그룹이 동일한 모집단에서 추출되었다는 가정을 검정하기 위한 비모수적 통계절차이다. 모든 데이터 그룹을 취합한 후 개별 값을 오름차순으로 정렬하여, $z_{(1)}$, $z_{(2)}$, $\cdots$, $z_{(L)}$ 로 표기한다. 이 때 동일한 수치의 데이터가 있다면 L은 전체 데이터 개수 n보다 작을 것이다. 이러한 순위는 ADK 검정통계량 계산에 사용된다. ADK 검정통계량은 다음과 같다.

$$ADK = \frac{n-1}{n^2(k-1)} \sum_{i=1}^{k} \left[\frac{1}{n_i} \sum_{j=1}^{L} \frac{h_j(nF_{ij}-n_iH_j)^2}{H_j(n-H_j)-nh_j/4} \right] \quad (1)$$

여기서, n_i는 각 뱃치의 시편 개수, $n=\sum_{i=1}^{k} n_i$, h_j는 취합된 표본 중 $z_{(j)}$와 동일한 값의 개수, H_j는 취합된 표본 중 $z_{(j)}$보다 작은 값의 개수와 $z_{(j)}$와 동일한 값의 개수의 ½의 합, F_{ij}는 i번째 그룹 중 $z_{(j)}$보다 작은 값의 개수와 $z_{(j)}$와 동일한 값의 개수의 ½의 합이다.

ADK 검정통계량이 다음 식으로 계산되는 임계값(ADC)보다 크다면, α의 오차확률로 각 데이터 그룹이 서로 다른 모집단에서 추출되었다고 결론내릴 수 있다. 반대의 경우에는 각 그룹이 동일 모집단에서 추출되었다는 가정이 성립하며, 뱃치 간 변동성이 무시할만하다고 결론내릴 수 있다.

$$ADC(\alpha) = 1 + \sigma_n \left(b_0 + \frac{b_1}{\sqrt{k-1}} + \frac{b_2}{k-1} \right) \quad (2)$$

여기서, 상수 b_0, b_1, b_2는 α에 따라 다르며, 예를 들어 다음과 같다.

Table 1. Interpolation Coefficients

α	b_0	b_1	b_2
0.05	1.645	0.678	−0.362
0.025	1.960	1.149	−0.391
0.01	2.326	1.822	−0.396

뱃치 간 변동성에 대한 ADK 검정을 통과하지 못한다면, CMH-17의 요구조건을 충족시키는 기저값 도출방법은 ANOVA 방법이 유일하다.

3.3 분포함수에 대한 적합도 평가

꼬리 영역에서의 차이에 민감한 Anderson Darling 검정통계량을 이용하여 각 분포에 대한 적합도를 조사한다. Anderson Darling 검정은 주어진 데이터의 누적분포함수를 예상 누적분포함수와 비교한다. 우선 데이터를 예상 분포에 대한 공통적인 표현으로 변환한다. 예를 들어, 정규분포에 대해서는 0의 평균과 1의 표준편차를 갖도록 데이터를 정규화한다. Anderson Darling 검정통계량에 근거하여 관측유의수준(OSL)을 계산한다. OSL이 0.05보다 작거나 같으면, 해당 데이터가 예상되는 분포에서 추출되었다는 가정은 최대 5%의 오차 확률로 기각된다.

3.3.1 정규분포

정규분포 모집단의 평균과 표준편차는 표본평균 $\bar{x}$과 표본표준편차 s를 이용하여 다음과 같이 추정한다.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (3)$$

$$s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (4)$$

정규분포에 대한 적합도를 평가하기 위해, 데이터에 가장 적합한 정규 누적분포함수와 데이터의 실제 누적분포함수를 비교한다. $z_{(i)}$를 다음과 같이 정의하면,

$$z_{(i)} = \frac{x_{(i)} - \bar{x}}{s}, \quad \text{for } i = 1, \ldots, n \qquad (5)$$

여기서, $x_{(i)}$는 i번째 표본값이다.

Anderson Darling 검정통계량은 다음과 같다.

$$AD = \sum_{i=1}^{n} \frac{1-2i}{n}\left[\begin{array}{c}\ln\{F_0(z_{(i)})\}+ \\ \ln\{1-F_0(z_{(n+1-i)})\}\end{array}\right] - n \qquad (6)$$

여기서 F_0는 표준정규분포의 누적분포함수이며, 관측유의수준은 다음과 같다.

$$OSL = \frac{1}{1+\exp[-0.48+0.78\ln(AD^*)+4.58AD^*]} \qquad (7)$$

여기서, $AD^* = \left(1+\frac{4}{n}-\frac{25}{n^2}\right)AD$이다.

OSL ≤ 0.05 라면, 5%의 오차 확률로 모집단이 정규분포를 이루지 않는다고 결론내릴 수 있다. OSL > 0.05 라면, 모집단이 정규분포를 이룬다는 가설은 성립된다.

3.3.2 Weibull 분포

2-모수 Weibull 분포의 모집단에 대한 기저값을 계산하기 위해서는, 모집단의 형상(shape) 매개변수와 척도(scale) 매개변수의 추정값을 계산해야 하며, 이를 위해 최대우도법(maximum likelihood method)을 적용한다.

형상 매개변수와 척도 매개변수의 최대우도 추정값을 각각 $\hat{\beta}$와 $\hat{\alpha}$로 표기한다. 추정값은 다음 연립방정식의 해이다.

$$\hat{\alpha}\hat{\beta}n - \frac{\hat{\beta}}{\hat{\alpha}^{\hat{\beta}-1}}\sum_{i=1}^{n} x_i^{\hat{\beta}} = 0 \qquad (8)$$

$$\frac{n}{\hat{\beta}} - n\ln\hat{\alpha} + \sum_{i=1}^{n}\ln x_i - \sum_{i=1}^{n}\left[\frac{x_i}{\hat{\alpha}}\right]^{\hat{\beta}}(\ln x_i - \ln\hat{\alpha}) = 0 \qquad (9)$$

Weibull 분포에 대한 적합도를 평가하기 위해, 데이터에 가장 적합한 Weibull 누적분포함수와 데이터의 실제 누적분포함수를 비교한다. 다음과 같이 $z_{(i)}$를 정의하면,

$$z_{(i)} = \left[x_i/\hat{\alpha}\right]^{\hat{\beta}}, \quad \text{for } i = 1, \ldots, n \qquad (10)$$

Anderson Darling 검정통계량과 관측유의수준은 각각 다음과 같다.

$$AD = \sum_{i=1}^{n} \frac{1-2i}{n}\left[\begin{array}{c}\ln\{1-\exp(-z_{(i)})\} \\ -z_{(n+1-i)}\end{array}\right] - n \qquad (11)$$

$$OSL = \left\{1+e^{-0.10+1.24\ln(AD^*)+4.48AD^*}\right\}^{-1} \qquad (12)$$

$$AD^* = \left(1+0.2/\sqrt{n}\right)AD \qquad (13)$$

OSL ≤ 0.05 라면, 5%의 오차 확률로 모집단이 2-모수 Weibull 분포가 아니라는 결론을 내릴 수 있다. OSL > 0.05 라면, 모집단이 2-모수 Weibull 분포를 갖는다는 가설이 성립된다.

3.3.3 로그정규분포

어떤 데이터가 로그정규분포를 이루면, 그것의 로그값은 정규분포를 이룬다. 로그정규분포에 대한 적합도를 평가하기 위해, 주어진 데이터의 로그값에 대해 정규성을 위한 Anderson Darling 검정을 수행한다. 자연로그를 이용하여 Eq. 5를 다음 식으로 대체한다.

$$z_{(i)} = \frac{\ln(x_{(i)}) - \overline{x}_L}{s_L}, \quad \text{for } i = 1, \ldots, n \quad (14)$$

여기서, $x_{(i)}$는 i번째로 작은 표본값이고, $\overline{x}_L$과 s_L은 $\ln(x_{(i)})$의 평균과 표준편차이다.

Anderson Darling 통계량과 관측유의수준은 각각 Eq. 6과 Eq. 7을 이용하여 계산된다.

OSL ≤ 0.05 라면, 5%의 오차 확률로 모집단이 로그정규분포가 아니라는 결론을 내릴 수 있다. OSL > 0.05 라면, 모집단이 로그정규분포를 갖는다는 가설이 성립된다.

3.3.4 분포함수 모델의 선택

다음과 같은 이유로 인해 정규분포가 선호된다.

- 복합재 강도 데이터의 표본이 커지면 정규분포를 이루는 경향이 있다고 알려져 있다.
- 많은 양의 데이터가 없다면, 합당한 수준의 통계적 검증력을 가지고 정규분포, Weibull 분포, 또는 로그정규분포를 따르는 데이터집합 사이의 차이를 구별하는 것이 불가능하다.
- MNR 방법과 분산동질성 검정을 적용할 때 이미 정규분포 거동을 가정하고 있다.

따라서, 반대되는 강력한 증거가 없다면, 정규분포를 사용하는 것이 낫다. 일반적으로, 정규분포 OSL이 0.05보다 크다면, 정규분포를 이용한다. 하지만, 정규분포 OSL이 0.05보다 크다 해도, 다음과 같은 이유로 다른 모델을 선택할 수도 있다.

해당 물성에 대한 다른 환경조건들에 대해 정규분포 OSL이 0.05보다 작아 다른 모델을 적용할 수밖에 없었다면, 고려중인 환경조건에 대해서도 일관성을 위해서 다른 환경조건들에 대해 적용된 모델을 선택할 수 있다. 단, 고려중인 환경조건의 해당 모델에 대한 OSL이 0.05보다 커야 한다.

다른 모델에 대한 OSL이 정규분포 OSL보다 상당히 크다면, 다른 모델을 고려할 수 있다. 이러한 OSL의 차이는 일반적으로 큰 데이터집합에 대해서만 중요하다.

정규분포 OSL이 0.05보다 작다면, OSL의 크기에 따라 또는 해당 물성의 다른 환경조건과의 일관성을 고려하여 다른 모델을 선택해야 한다. 정규분포 OSL이 0.05보다 작고, Weibull 분포 또는 로그정규분포 중 하나가 0.05보다 큰 OSL을 갖는 경우에는 Weibull 분포 또는 로그정규분포 중 하나를 사용할 수 있다. 세 모델 중 주어진 데이터에 적합한 모델이 발견되지 않으면(모든 OSL 값이 0.05보다 작거나 같은 경우), 비모수적 방법으로 기저값을 계산해야 한다.

4. 결 언

복합재 허용치 계산을 위한 통계분석기법 중 각 환경조건 데이터를 개별적으로 분석하는 점추정 방법과 분석 과정에서 주의해야 할 점에 대해 알아보았다. 뱃치 간 변동성이 심각한 환경조건 데이터집합을 제외한 후 남아있는 환경조건의 개수가 2개 이상이 되지 못하거나, 환경조건 간 파단 모드의 일관성이 없거나, 정규성 조건을 만족하지 못하는 경우와 같이 합동 방법을 적용할 수 없는 상황이라면, 점추정 방법을 적용하여 각 환경조건 데이터를 개별적으로 분석해야 한다. 점추정 방법에서는, 환경조건 별로 적용 분포함수 모델이 다를 수 있기 때문에, 기저값과 평균값의 경향이 환경조건 별로 바뀔 수 있으며, 복합재 구조물 설계에 필요한 하중향상계수 및 환경보정계수를 현실적인 값으로 얻기가 상당히 어렵다. 또한, 점추정 방법으로 계산한 B-기저값이 평균값의 90% 이상인 경우는, 변동계수가 비정상적으로 낮고 보수적이지 않을 수 있다는 것을 시사하므로 주의가 요구된다.

참고문헌

[1] SAE International, CMH-17 Rev.G, Composite Materials Handbook - Vol.1. Polymer Matrix Composites Guidelines for Characterization of Structural Materials, 2012.
[2] Tomblin, J., Ng, Y., and Raju, K. “Material Qualification and Equivalency for Polymer Matrix Composite Material Systems: Updated Procedure,” FAA report DOT/FAA/AR-03/19, 2003.