

수박의 내부품질 예측모델 개발

Development of Prediction Model for Internal Quality of Watermelon

최동수*	최규홍*	이강진*	이영희*	김만수**
정회원	정회원	정회원	정회원	정회원
D. S. Choi	K. H. Choi	K. J. Lee	Y. H. Lee	M. S. Kim

1. 서론

수박 선별은 1차로 생산자가 무게, 과형, 과피 색택 등 경험에 의한 주관적인 내부품질 판정을 하는 경우와 산지공판장 또는 가락동 도매시장에서 경매 전에 하역하면서 위와 유사한 기준과 방법으로 유통인에 의하여 선별된다. 따라서 유통비용에는 선별비가 포함되고, 경매장소에 따라 선별비의 차이가 많은데 수박의 경매물량이 가장 많은 가락동 도매시장 경우, 5톤 트럭(1000개) 1대당 하역비 7만원, 선별비가 7만원이 드는 것으로 조사되어있다. 이와 같은 인력에 의한 수박의 선별은 비용이 많이들 뿐만 아니라 선별의 객관성이 결여되어 소비자로부터 신뢰도 얻지 못하는 실정이다. 이러한 불합리적인 요소들을 제거하여 농가의 소득을 향상시키기 위해서는 객관적이고 비파괴적인 선별기술의 개발과 선별비용의 절감이 절실히 요구되고 있다.

외국에서는 수박을 비파괴적으로 선별하기 위한 연구가 오래 전부터 수행되어 최근에는 실용화 단계까지 와 있다. 하지만 국내에서는 아직까지 수박에 대한 비파괴적인 판정기술이 개발되어 있지 않으나 일부 연구자들에 연구가 수행 중에 있다.

수박의 내부품질인자로 속도, 공동, 황대, 피수박 등이 있는데, 최 등(1999, 2000)은 인공 신경망을 이용하여 내부품질 예측모델을 개발하였다. 4가지 품질인자에 대한 인공신경망모델 검증의 정확도는 각각 93, 80, 79, 88%라고 보고하였다.

본 연구에서는 수박의 음파특성을 근거로 하여 내부품질요인별 다중회귀모델을 이용하여 내부품질 예측모델을 개발하고자 하였다.

2. 재료 및 방법

가. 회귀분석법을 이용한 품질판정 모델개발

수박의 음파특성 분석자료를 이용하여 내부품질을 예측할 수 있는 모델을 개발하기 위하여 식(1)과 같은 다중회귀모델(Multiple Linear Regression model, MLR model)을 유도하였다.

* 농업기계화연구소
* 충남대학교 농과대학 농업기계공학과

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (1)$$

여기서, y = 속도, 공동, 황대, 피수박의 품질지수

a = 절편

$b_1 \dots b_n$ = 회귀계수

$x_1 \dots x_n$ = 내부품질 특징인자

속도 회귀모델에 사용된 종속변수는 과육색, 밀도, 경도, 씨의 착색정도 등을 고려하여 완전 미숙에서 과숙까지 속도의 정도에 따라 70~140의 수치를 부여하여 속도지수로 하였으며 독립변수는 주파수 스펙트럼 에너지비 (0~550)/(850~2500Hz), (0~850)/(850~2500Hz), 제1, 2 피크 주파수 등으로 하였다.

공동의 경우 종속변수는 공동의 유무에 따라, 1과 0의 수치를 부여하여 공동지수로 하였으며, 독립변수는 주파수 스펙트럼 에너지비 (0~550)/(850~2500Hz), (0~850)/(850~2500Hz), 제1, 2, 3 피크 주파수 등이었다.

황대의 경우 황대가 없을 때 0, 가장 심할 경우 5의 수치를 부여한 황대지수를 종속변수로 사용하였으며, 독립변수는 무게, 주파수 스펙트럼 에너지비 (0~550)/(850~2500Hz), 정규화된 주파수 스펙트럼에서 600~900Hz 대역의 에너지, 이 대역에서 최대 피크 주파수와 매그니튜드 값, 전 대역에서 최대 매그니튜드 값 등이었다.

피수박의 경우는 피수박 정도에 따라 정상과는 0, 가장 심할 때 5의 수치를 부여한 피수박지수를 종속변수로 사용하였으며, 독립변수는 무게, 주파수 스펙트럼 에너지비 (0~550)/(850~2500Hz), 제1, 2 피크 주파수, 시간파형의 제 1, 2, 최대 피크의 진폭값등 이었다.

여기서 각 품질지수에 대한 예측 모델 개발에 사용된 시료의 수는 속도 24개, 공동 53개, 황대 26개 및 피수박 36개 이었다.

개발된 품질예측 모델의 평가는 모델개발에 사용된 각 시료에 대한 측정값과 예측모델의 예측 값 사이의 오차에 대한 표준편차인 SEC(Standard Error of Calibration), SEP(Standard Error of Prediction), 결정계수 R^2 를 이용하여 평가하였으며, SEC, SEP, Bias 는 각각 식(2), 식(3), 식(4)와 같다.

$$SEC = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n - p - 1}} \quad (2)$$

$$SEP = \sqrt{\frac{\sum_{i=1}^n [(\hat{y}_i - y_i) - Bias]^2}{n - 1}} \quad (3)$$

$$Bias = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (4)$$

여기서, \hat{y}_i = 품질예측 모델에 의한 시료의 예측된 품질지수

y_i = 실험에 의해 평가된 품질지수

n = 시료의 수

p = 회귀모델 작성에 사용된 변수의 수

3. 결과 및 고찰

가. 속도 예측모델

수박의 속도를 예측하기 위해 개발된 예측모델은 식(5)와 같으며 그림 1에 나타낸 바와 같이 개발된 회귀식의 결정계수 R^2 과 SEC는 각각 0.6344, 14.212로 비교적 좋은 결과를 얻었다. 이 모델식을 이용하여 미지의 시료 17개의 속도를 예측한 결과 결정계수 R^2 과 SEP는 각각 0.3501, 19.3012 이었다. 속도지수 100(전문가에 의한 미숙과 적숙의 기준지수)을 기준으로 하였을 경우 모델의 미숙과 적숙판정의 정확도는 83.3%, 과숙 판정 정확도는 96.0% 이었다. 미지의 시료17개에 대한 미숙과 적숙 판정의 정확도는 76.5%, 과숙 판정의 정확도는 88.2% 이었다. 개발된 다중회귀모델에서도 신경망의 경우와는 달리 판정정확도가 100%가 되지 않는 것은 회귀모델의 한계이며 이것이 100%가 되기 위해서는 모든 측정들이 회귀식에 정확히 일치하여야 할 것이다.

$$\hat{y} = 18.4144 \times E_1 - 15.3916 \times E_2 + 1.0368 \times f_1 - 1.6162 \times f_2 + 299.6503 \quad (5)$$

여기서, \hat{y} = 속도지수 예측 값(70~140)

E_1 = 주파수대역의 에너지비((0~550)/(850~2500Hz))

E_2 = 주파수대역의 에너지비((0~850)/(850~2500Hz))

f_i = 주파수 스펙트럼에서 i번째 피크 주파수, Hz

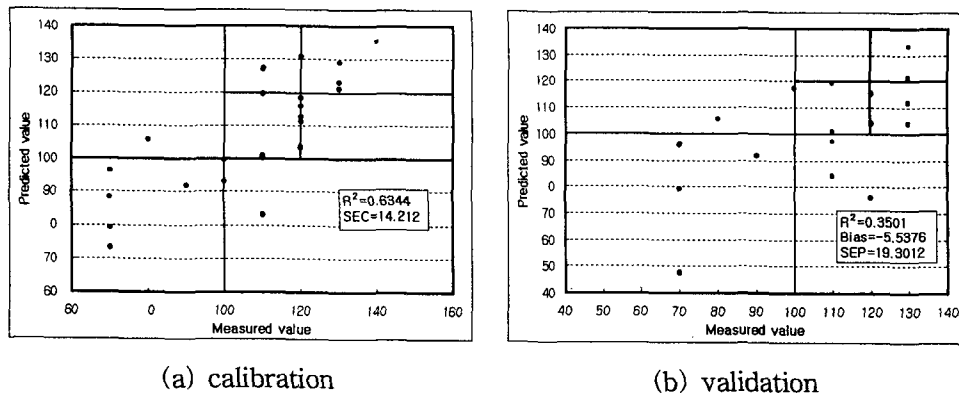


Fig. 1 Relationship between the measured quality index value and the predicted value by MLR for ripeness of the watermelon.

나. 공동 예측모델

공동과 예측을 위한 다중회귀모델은 식(6)와 같으며 그림 2에 나타낸 바와 같이 개발된 회귀식의 결정계수 R^2 , SEC, SEP는 각각 0.5783, 0.3031, 0.3342 이었다. 결정계수는 높지 않았으나 SEC와 SEP값의 차가 매우 적은 것을 보면 공동예측모델에서 선택된 독립변수는 합리적인 것으로 생각되었다. 공동에 대한 품질지수 0.3을 문턱값으로 설정하여 공동의 유무를 판정한 결과 예측모델의 정확도는 94.3%이었으며, 미지시료의 검증 정확도는 92.5%로 높은 결과를 얻었는바, 다중회귀모델에 의하여 수박의 공동유무를 판정한다고 하더라도 비교적 정확하게 판정할 수 있을 것으로 판단되었다.

$$\hat{y} = -0.01644 \times f_1 + 0.00891 \times f_3 + 0.14842 \times E_1 - 0.03054 \times E_2 - 0.00038 \times f_2 + 0.04321 \quad (6)$$

여기서, \hat{y} = 공동지수 예측 값
 f_i = 주파수 스펙트럼에서 i번째 피크 주파수, Hz
 E_1 = 주파수대역의 에너지비((0~550)/(850~2500Hz))
 E_2 = 주파수대역의 에너지비((0~850)/(850~2500Hz))

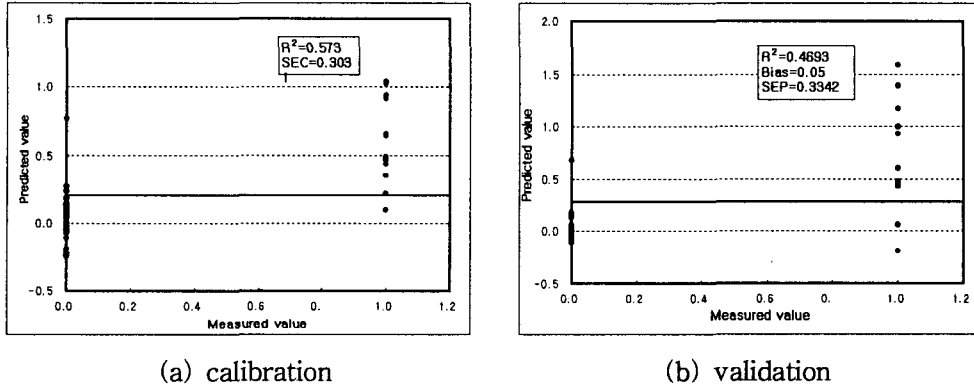


Fig. 2 Relationship between the measured quality index value and the predicted value by MLR for inside-cavity of the watermelon.

다. 황대 예측모델

개발된 황대 예측모델은 식(7)과 같으며, 그림 3에서 보는바와 같이 개발된 모델의 결정 계수 R^2 , SEC, SEP는 각각 0.8262, 0.6333, 1.0588 이었다. 이 모델의 결정계수는 상당히 높게 나타났으며 SEC와 SEP의 차도 비교적 적게 나타났으므로 황대과의 판별정확도도 높을 것으로 판단되었다.

황대의 품질 지수 1.5로 문턱값을 설정하여 황대의 유무를 판정한 결과 그림 3에서 보는 바와 같이 모델의 정확도는 88%이었으며 미시자료 14개에 대하여 검증하였던 결과 정확도는 93%로 높은 결과를 얻었다. 이와 같이 검증결과에서도 판별정확도가 높은 것은 황대과의 경우 속도나 공동의 경우와는 달리 음파 특성에서 에너지비, 최대 매그니튜드값, 주파수대역에서의 매그니튜드값 및 피크주파수 등에서 특징점이 나타나므로 이들을 독립변수로 이용할 수 있었기 때문인 것으로 판단 되었다.

$$\hat{y} = -0.05123 \times W + 0.27294 \times E_1 + 0.06901 \times E_2 + 0.00771 \times M_1 + 1.33417 \times M_2 + 0.00478 \times f - 6.01007 \quad (7)$$

여기서, \hat{y} = 황대지수 예측 값
 W = 공시재료의 질량, kg
 E_1 = 주파수대역의 에너지비((0~550)/(850~2500Hz))
 E_2 = 주파수대역의 정규화된 에너지(600~900Hz)
 M_1 = 주파수 스펙트럼에서 최대 매그니튜드 값
 M_2 = 600~900Hz 주파수대역에서 최대 피크 주파수의 매그니튜드 값
 f = 600~900Hz 주파수대역에서 피크 주파수, Hz

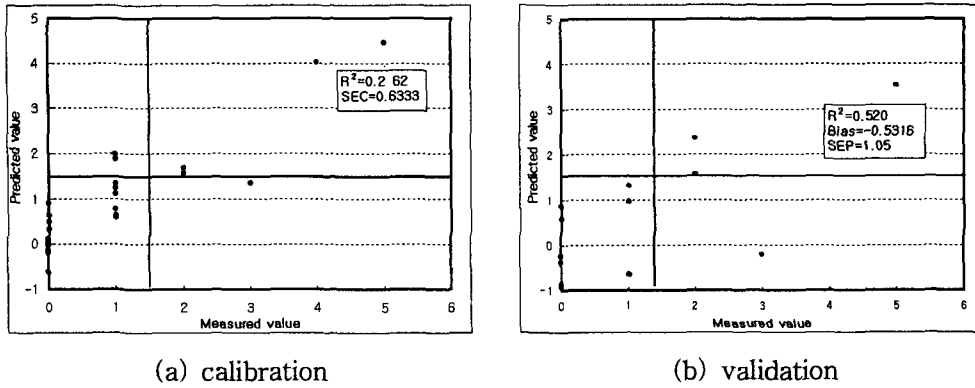


Fig. 3 Relationship between the measured quality index value and the predicted value by MLR for yellow belt-watermelon.

라. 피수박 예측모델

개발된 피수박 예측 모델은 식(8)과 같으며, 그림 4에서 보는바와 같이 개발된 모델의 결정계수 R^2 , SEC, SEP는 각각 0.8007, 0.8978, 2.2098이었다. 이 모델의 경우도 황대에서의 유사하게 상당히 높은 결정계수와 SEC와 SEP값의 차이도 비교적 적은 것으로 분석되었다.

$$\hat{y} = 0.01831 \times W + 0.56516 \times E - 0.04067 \times f_1 + 0.01947 \times f_2 - 0.33966 \times A_1 - 0.95888 \times A_2 - 2.11214 \times A_3 + 3.8825 \quad (8)$$

여기서, \hat{y} = 피수박지수 예측 값
 W = 공시재료의 질량, kg
 E = 주파수대역의 에너지비((0~550)/(850~2500Hz))
 f_i = 주파수 스펙트럼에서 i번째 피크 주파수, Hz
 A_1, A_2 = 시간파형에서 제 1, 2 피크의 진폭값, V
 A_3 = 시간파형에서 최대 진폭값, V

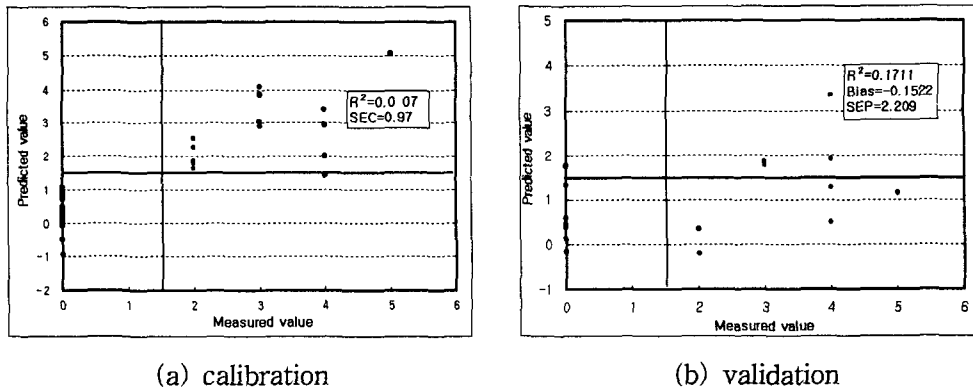


Fig. 4 Relationship between the measured quality index value and the predicted value by MLR for blood flesh-watermelon.

피수박의 품질지수 1.5로 문턱값을 설정하여 피수박의 유무를 판정한 결과 그림 4에서와 같이 모델의 정확도는 97%로 높은 결과를 얻었으나 미지시료 20개에 대하여 검증한 결과 정확도는 70%로 매우 낮게 나타났다. 이러한 현상은 수박 개체간의 차이가 큰데서 기인

된 것으로 생각되었으며 시료의 수를 좀더 많이 한다면 모델의 판정 정확도에서와 같이 좋은 결과를 얻을 수 있을 것이고 따라서 피수박의 판정은 개발된 다중회귀모델로 충분히 가능하다고 판단되었다.

4. 요약 및 결론

본 연구는 다중회귀모델을 이용하여 수박의 내부품질 예측모델을 개발하였으며, 그 결과를 요약하면 다음과 같다.

1. MLR 모델을 이용하여 품질인자들에 대해 예측 모델을 개발한 결과 속도의 경우 예측모델의 결정계수 R^2 과 SEC는 각각 0.6344, 14.212이며, 속도지수 100을 기준으로 미숙과 적숙을 판정한 결과 모델의 정확도는 83.3%로 비교적 좋은 결과를 얻었다. 이 모델식을 이용하여 미지의 시료 17개의 속도를 예측한 결과 결정계수 R^2 과 SEP는 각각 0.3501, 19.3012 이었으며, 미숙과 적숙 판정의 정확도는 76.5%, 과숙과에 대한 판정 정확도는 88.2% 이었다.
2. MLR 모델을 이용해 개발된 공동예측 모델은 결정계수 R^2 과 SEC는 각각 0.5783, 0.3031 이었으며, 공동의 품질지수 0.3을 문턱값으로 설정하여 공동의 유무를 판정한 결과 예측모델의 정확도는 94.3%이었으며, 미지시료에 의한 검증 결과 판정 정확도는 92.5%로 높게 나타났다.
3. MLR 모델을 이용해 개발된 황대예측 모델은 결정계수 R^2 과 SEC는 각각 0.8262, 0.6333 이었으며, 황대의 품질지수 1.5를 문턱값으로 설정하여 황대의 유무를 판정한 결과 모델의 정확도는 88%이었으며 미지시료에 의한 검증한 결과 판정 정확도는 93%로 비교적 높게 나타났다.
4. MLR 모델을 이용해 개발된 피수박 예측 모델은 결정계수 R^2 과 SEC는 각각 0.8007, 0.8978이었으며, 피수박의 품질지수 1.5를 문턱값으로 설정하여 피수박의 유무를 판정한 결과 모델의 정확도는 97%로 높은 결과를 얻었으나 미지시료로 검증한 결과 판정 정확도는 70%로 매우 낮은 값으로 나타났다.

5. 참고문헌

1. 성내경. 1998. Windows용 SAS를 이용한 통계적 자료분석. 자유아카데미.
2. 김만수, 최동수, 이영희, 조영길. 1998. 수박의 음향특성에 관한 연구. 한국농업기계학회지 23(1); 57-66
3. 최동수, 최규홍, 이영희, 이강진, 윤진하, 김만수. 1999. 인공신경망을 이용한 수박의 음향 특성 분석(I)-속도 및 공동. 한국농업기계학회 동계학술대회 논문집4(1): 652-660.
4. 최동수, 최규홍, 이강진, 박원규, 김만수, 최동철, 황영. 2000. 인공신경망을 이용한 수박의 음향특성 분석(II)-황대 및 피수박. 한국농업기계학회 동계학술대회 논문집 5(1): 422-429.