

Random forest의 중요도 지수로 결정한 우리나라 벼의 기상학적 수량결정요인 분석

김준환*, 상완규, 신평, 조현숙, 서명철
농촌진흥청 국립식량과학원 작물재배생리과

Meteorological Factors of Rice Yield using Variable Importance of Random Forest

Junhwan Kim*, Wangyu Sang, Pyeong Shin, Hyeounsuk Cho and Myungchul Seo
*Crop physiology and production, National Institute of Crop Science, Rural Development Administration, 181,
Hyeoksin-ro, Iseo-myeon, Wanju-Gun, Jeollabuk-do 55365, Republic of Korea*

Random forest는 기계학습의 하나로 분류와 예측에 사용할 수 있는데 본 연구에서는 벼작황과 기상과의 관계를 random forest로 이용하였을 때 설명이 가능한지 또한 이로부터 얻어지는 중요도지수(random forest)가 실제의 경험적인 관찰과 일치하는지 검토하는데 있다. 1981년부터 2015년까지의 통계청 수량자료와 기상청 관측기상자료를 각각 평균하여 이용하였으며 기상자료는 순별로 정리하여 사용하였다. Random forest를 이용한 예측은 전반적으로 실제의 수량 추세를 시계열상에서 잘 반영하여 relative root mean square (%) 는 2.3%로 상당히 높았다(Fig. 1). Random forest 에서 얻어진 상위 중요도지수(Variable importance)는 년도, 10월 하순 최저 및 평균온도 5월 중순 최저온도 6월 하순 평균온도 순이었다(Fig. 2). 년도는 국내 벼 재배기술의 기술적 진보를 보여주는 것으로 중요 변수로 포함된 것은 타당하다고 할 수 있다. 10월의 최저 및 평균온도는 벼의 등숙 가능 기간을 결정하는 변수로서 전국적인 규모에서는 수량에 영향을 주는 것도 가능할 것으로 보인다. 5월 중순의 최저온도는 육묘와 이앙 후 벼의 활착과 관련이 있으나 이 시기가 수량결정 상위에 위치할 정도로 중요한 시기는 아니다. 6월 하순은 시기상으로 분얼기로서 이시기의 평균온도가 의미가 있을 수가 있다. 전반적으로 의미가 있는 것도 있으나 그렇지 않은 것도 있어 이 방법에서 얻어진 변수를 통해 생육에 중요한 시기와 요소를 결정하는 것은 더 많은 연구가 필요할 것으로 보인다.

* Correspondence to : sfumato@korea.kr

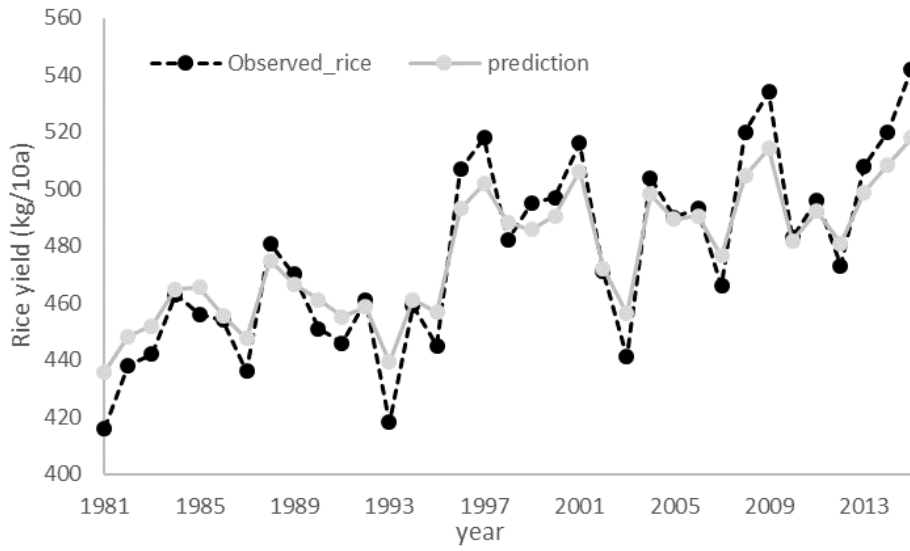


Fig. 1. 1981년부터 2015년까지의 실제수량과 예측수량

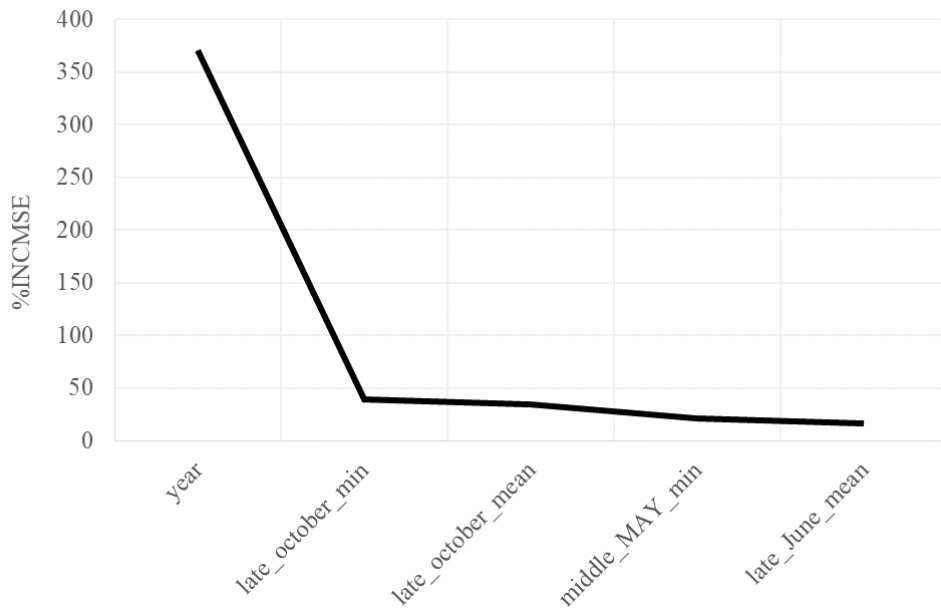


Fig. 2. random forest에서 얻어진 상위 중요도지수