Basic information

1. Subject: weather - US corn, soybeans (rainfed portion)

2. Duration: 1990 - 2017, May - August, weekly, biweekly, monthly

3. Location: Iowa (IA), Illinois (IL), Indiana (IN), Minnesota (MN), total 16 agricultural regions, 169 counties

4. Factors: total rainfall (Rain), average temperature (Tavg), maximum temperature difference (Tdiff)

5. Data source: PRISM, USDA-NASS

Overall results

1. Maize/soybean - growing season (especially: late development/early reproduction) - Rain/Tavg - important

- Maize - July/August - Tavg/Tdiff - Important

- Maize - June/July - Rain - Important

- Soybeans - June - Tavg - Important

- Soybeans - Aug - Tdiff/Rain - Important

2. Tdiff improvement

- SVM - Corn - Weekly - No improvement

- MLR - Corn - Monthly - No improvement

- SVM - Soybeans - Monthly - No improvement

- Other -> Significant improvement

3. Model quality: SVM - weekly ≈ biweekly > monthly

4. Average yield: maize - 11950kg/ha, soybean - 3680kg/ha

5. Coefficient of variation (CV): maize - 12%, soybean - 11.3%

6. Anomalies: 1993 floods, 2003 drought (August/September), 2012 major drought

Specific results

1. Maize - June/July Rain, July/August Tavg/Tdiff > other factors

- May/August Rain - early vegetative stage/late reproductive stage - low impact

2. Soybeans - June Tavg (early reproductive stage), August Rain, July/August Tdiff > other factors

3. Maize - SVM - Tavg/Rain - lowest RMSE (591kg/ha, < 5%nRMSE), highest r,d

GAM next (Tavg/Rain/Tdiff - 792kg/ha), MLR worst (Tavg/Rain/Tdiff - 1065kg/ha)

4. Soybean - SVM - Tavg/Rain - lowest RMSE (205kg/ha, < 6%nRMSE), highest r,d

GAM next (Tavg/Rain/Tdiff - 274kg/ha), MLR worst (Tavg/Rain/Tdiff - 320kg/ha)

5. Overall data at state level,

1) Maize:

- IA - best - 472kg/ha (3.8% nRMSE)

- MN - second - 505kg/ha (4.2%nRMSE)

- IL - third - 660kg/ha (5.4%nRMSE)

- IN - Worst - 733kg/ha (6.8%nRMSE)

- d - all > 0.9, r - IA, MN, IL > 0.9

2) Soybeans:

- IA - best - 178kg/ha (4.8%nRMSE)

- IL - second - 197kg/ha (5.1%nRMSE)

- MN - third - 199kg/ha (5.7%nRMSE)

- IN - Worst - 219kg/ha (6.1%nRMSE)

- d - all > 0.9, r - IA: 0.9, MN: 0.89, IL: 0.86, IN: 0.84

6. Collinearity between factors

- June vs July Tavg - Positive correlation

- June - Tavg vs Rain - Negative correlation

7. The role of Tdiff

- Tdiff important factor in July/August (reproductive period)

- Tdiff too low - higher night time temperatures - increased respiration and longer reproductive period - reduced crop yield

- Tdiff too high - also reduces crop yield

Model analysis

1. Using: stepwise multiple linear regression (MLR), general additive model (GAM), support vector machine (SVM)

2. Results: SVM > GAM > MLR

3. Optimisation: improvement after adding Tdiff, but related to model and weather data time level (improvement/mostly the same)

4. Initialisation: counter-trending treatment (technological advances, evolution of species), use of linear detrending (or difference, non-linear regression); centred, scaled

5. MLR: y = b0 + SUM(bi \* xi) + ε, based on MSE minimum

6. GAM: y = b0 + SUM(fi \* xi) + ε, with f as MLR

7. SVM: y = ωx + b, tolerance ε (ε insensitive region), based on minimum bias in insensitive region; first add kernel function -> linearisation of data (in this paper: radial basis function, or polynomials, sigmoid functions)

- Strong modelling capability for non-linear functions and high latitude data

8. Assessment

Root mean square error: RMSE = SQRT(SUM(Ei - Ai)2 / n)

Normalised root mean square error: nRMSE = RMSE / mean(real\_data)

Correlation coefficient (r): [-1, 1] -> perfectly negative linear correlation - perfectly positive linear correlation

Consistency coefficient (d): [0, 1] -> perfect mismatch - perfect match

Precautions

1. Rain-fed: weather factors account for more than 75%

2. Cumulative effects: consider the effect of one adverse phase on the next, delayed

3. Collinearity: consider the interaction between weather factors

4. Other available factors: standardised precipitation evapotranspiration index (SPEI), spectral index, normalised difference vegetation index, enhanced vegetation index, soils, varieties, crop management systems, solar radiation, water vapour pressure

5. Other available methods: field observation/sampling, crop simulation modelling (area, density, row spacing, yield)

6. Other available models: mixed models, random forests, neural networks

7. Use of multiple packages: Caret, MGCV, KernLab

Summary

1. Separate analyses according to different time periods and lengths of time (growth period, reproductive period, weekly, monthly)

2. Analysis using different models with different parameters (MLR, GAM, SVM) to analyse the performance of each model

3. Removal of disturbances: rainfed areas, detrending, removal of emergencies, initial processing of data (centering, deflating)

4. Consider different combinations of factors (with and without Tdiff), and their importance

5. Evaluate models: RMSE, nRMSE, r, d

6. Consider collinearity between factors, cumulative effects

7. Attempt to give biological explanations

8. Try to use other methods, consider other factors, use other models, other software packages