Data Mining of Weather and Climatic Data to Improve Risk Prediction of Fusarium Head Blight

L. V. Madden, ***A. B. Kriss***, **P. A. Paul**
(Ohio State University)
X. Xu
(East Malling Research, UK)

Fusarium Head Blight: A Sporadic Disease

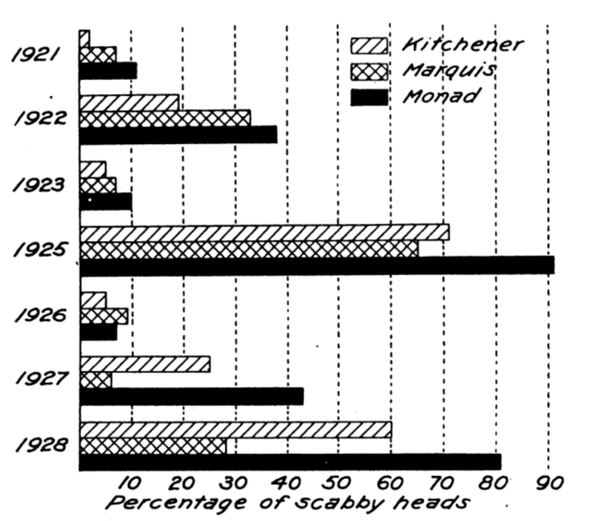
- Disease intensity (field severity and incidence) and mycotoxin (e.g., DON) varies considerably from location to location and from year to year
- In particular, the disease is <u>not rare</u>, and is <u>not so common</u> that a major epidemic occurs virtually every year
- There is considerable evidence from controlled experiments and empirical observations that epidemics depend on the environment (weather and/or climate)

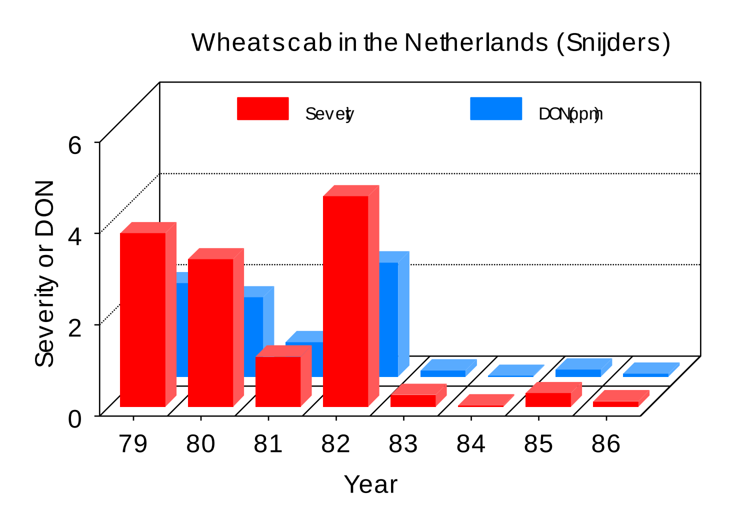
It is evident…that several variables strongly influence the percentage of infection. It seems probable that these variables may be meteorological conditions -- Chistristensen, Stakman, & Immer (1929)

There seems little doubt that the overriding limitation to FHB is moisture -- R. W. Stack (2000)

Christensen, Stakman, and Immer (1929)

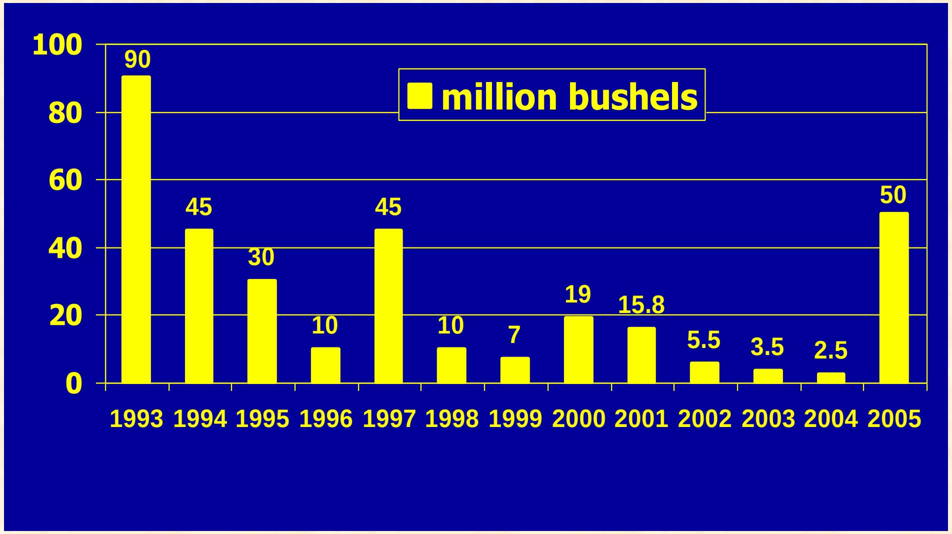
FHB: A long history





"During the period of 1927-1980, severe disease was reported in 1940, 1942, 1945, 1957, 1967, and 1980"

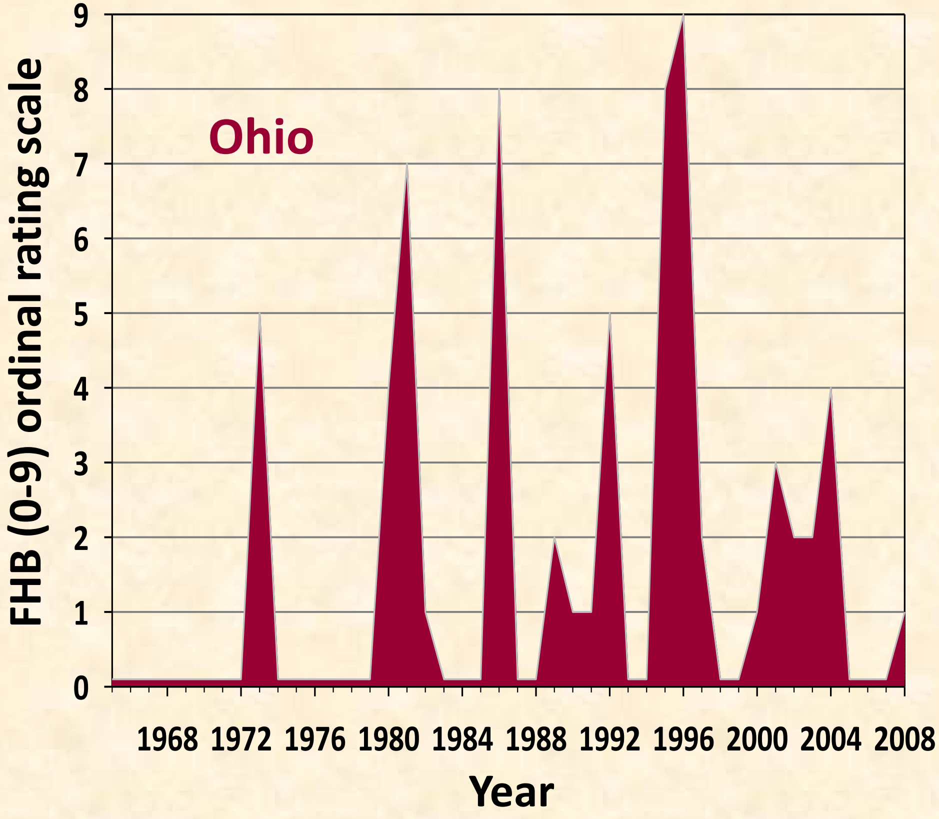
--J. C. Sutton ('82)

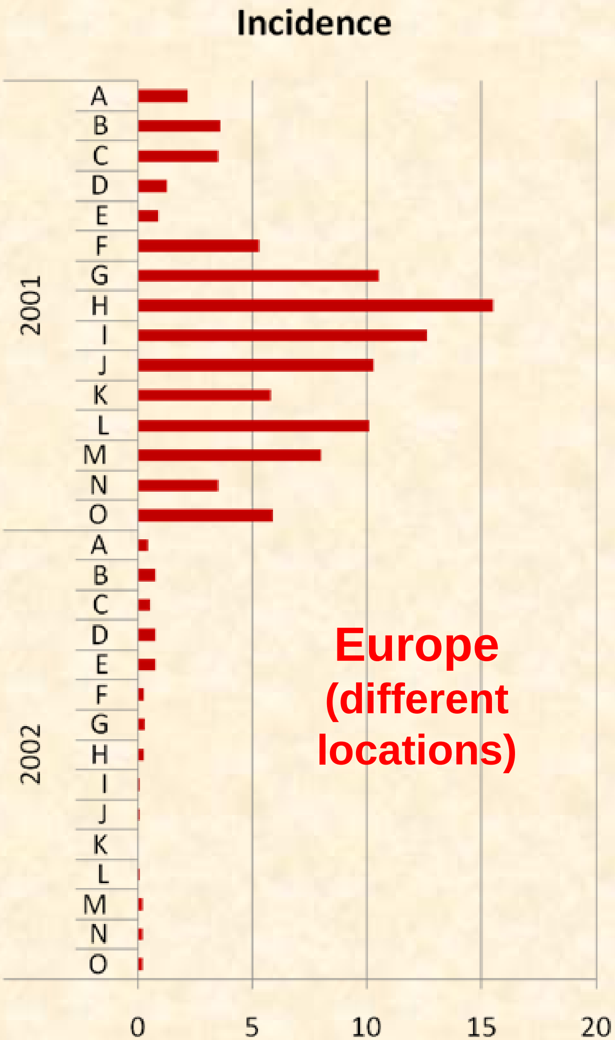


McMullen, M.

Temporal heterogeneity (within location)

Spatial heterogeneity





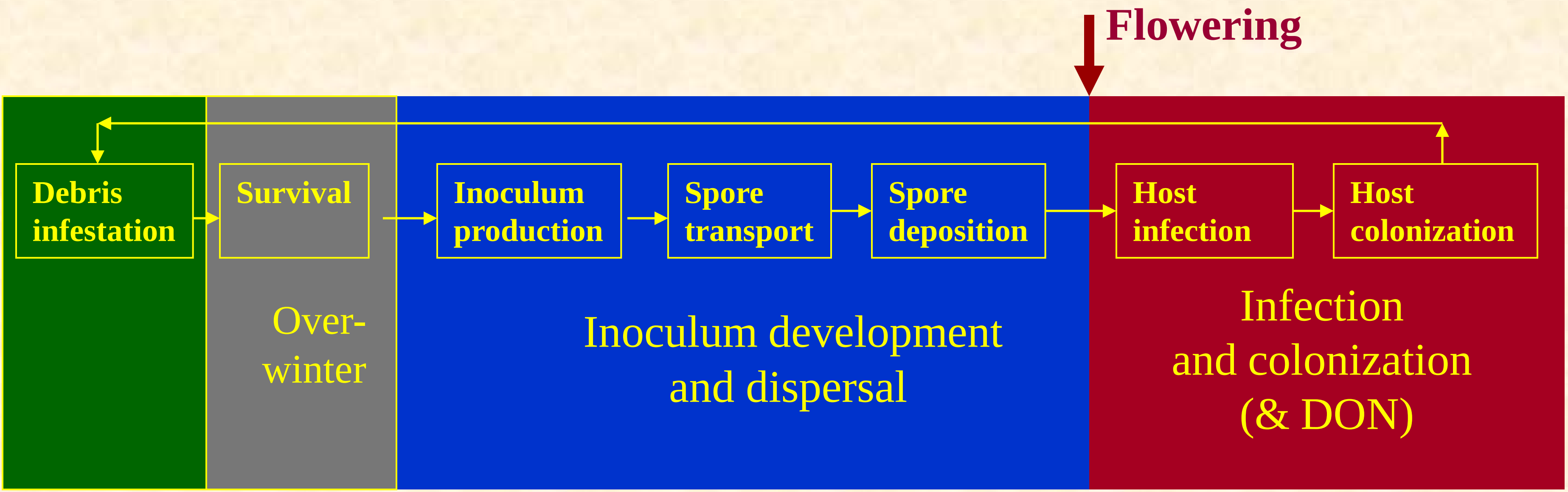
FHB and the environment

In principle, the occurrence of an epidemic may be due to:

Cumulative effects of the environment over different periods of time, coupled with effects of cropping practices, as well as unknowns

Time windows immediately preceding flowering up to harvest may be most informative/practical for predictions (and controls), but other windows can matter

What are the most important environmental variables and time windows?



‘Window Pane’ analysis:
Data mining for significant relationships in epidemiology

- Coakley et al. formalized the concept of **Window-Pane Analysis** in 1982, when they used it to develop models for stripe rust on winter wheat
  - Based on Coakley’s earlier work on climate-disease in late 1970s
- Concept:
  - **Determine time-window durations and starting (ending) times of windows when environmental variables are most highly correlated (associated) with disease intensity**
- Has been used for other diseases on wheat, potato, and rice
  - Recently refined and expanded to model disease-weather relationships for powdery mildew, yellow rust, and *Septoria tritici* epidemics on winter wheat, and now **FHB**
    - **Kriss, Paul, & Madden. 2010. *Phytopathology* 100: 784-797.**

FHB data sets

USA

- 4 locations (OH, IN, KS, ND), with many years (23-44) per location
- Environment: nearby weather stations
  - Long periods (data for *entire* year)
- Separate analysis/location
- Response variable: ordinal rating *or* FHB Index
- Analysis: Spearman rank correlation, …other…
- Kriss et al. *Phytopathology* 100: 784-797 (analysis)

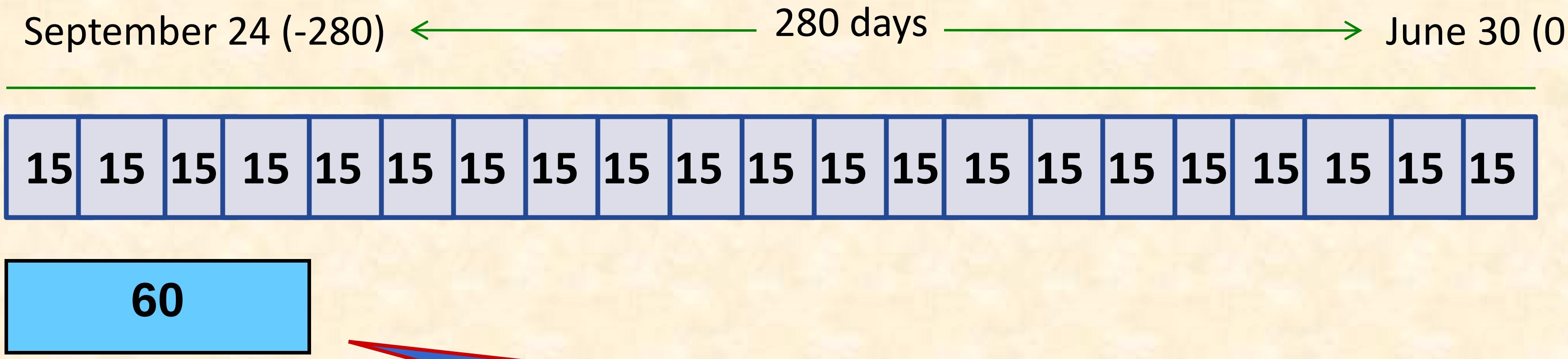
Europe

- 202 location-years; 4 countries (England, Ireland, Italy, Hungary), with up to 4 years/country
- Within-field micro-environment measurements
  - Short periods (*last* part of season)
- Pooled analysis
- Response: FHB Incidence, fungal biomass, *and* toxin
- Analysis: Spearman rank correlation, …other…
- Xu et al. *Phytopathology* 98: 69-78 (data collection)

‘Window Pane’ analysis

- Identify ‘summary’ environmental variables of potential interest, such as:
  - Average daily relative humidity; Hours with relative humidity > 80%; Average daily temperature; Total precipitation; Rain intensity; Hours of high relative humidity (e.g., > 80%) *and* temperature between 15 and 30 C; …
- Calculate the ‘summary’ environmental values for windows of different durations and starting (or ending) times
  - Example durations:
    - 5, 10, 15, 30, 60, … 280 days
  - Example starting times (if *t* = 0 is end of season and -280 is start of the winter-wheat season [previous year])
    - 0, -1, -2, -3, …. -280 days
- Calculate the relevant statistic (or statistics) for each window relating the environmental summary and biological variable
- *New* (Kriss et al.; 2010): correct for the multiple-testing problem in performing global and local tests of significance

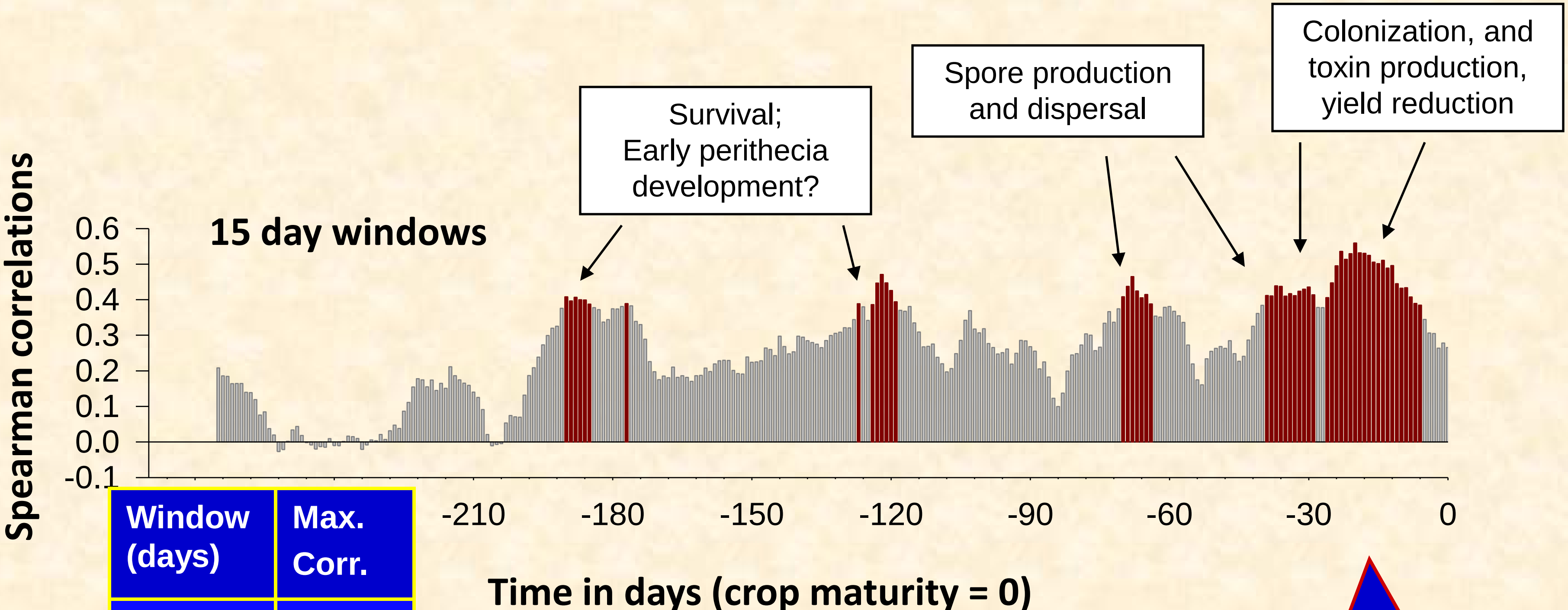
Window-Pane Construction: Example



Both shorter and longer windows are considered. Example: 60-days

Determine each environmental summary variable for each window size and time, and relate to disease using appropriate statistic(s)

Average Relative Humidity (ARH) and FHB Ordinal Ratings in Ohio

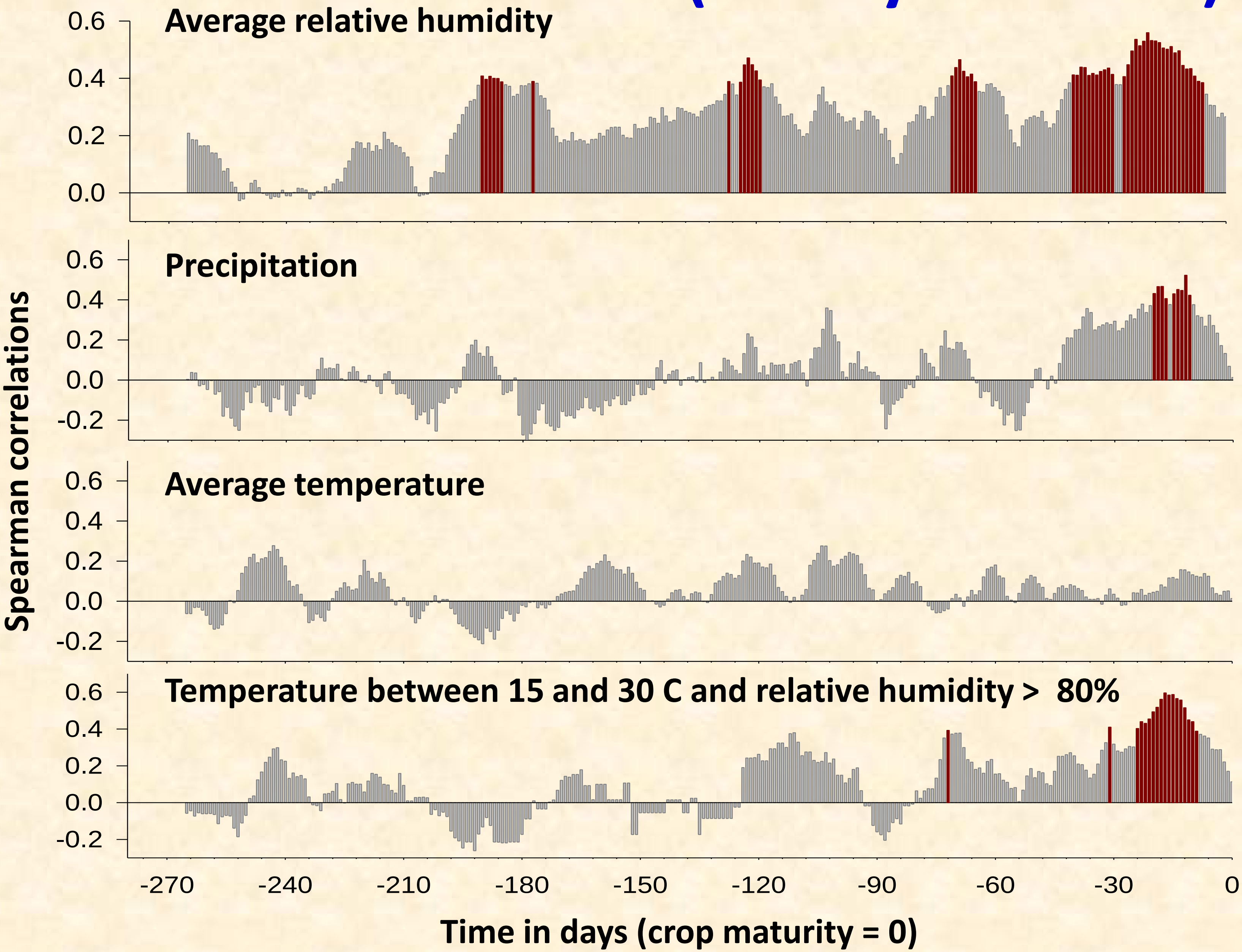


Window (days)	Max. Corr.
10	0.554
15	*0.559*
30	0.532
60	0.521
90	0.491
120	0.481

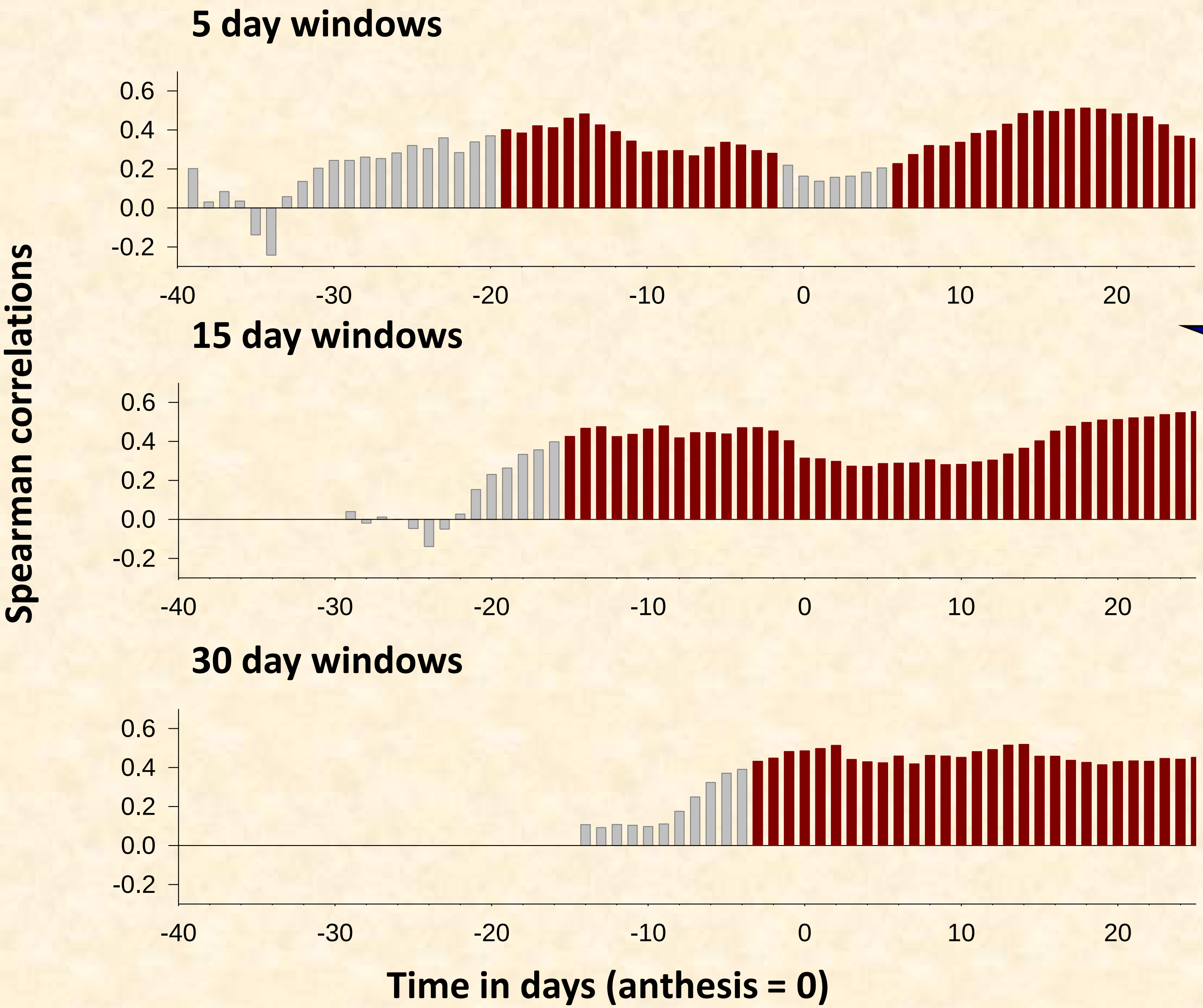
General trend of decreasing correlations with increasing window length

Each vertical bar is a correlation between ARH and FHB (dark red=significant) for a window "beginning" at identified time

Ohio (15-day windows)

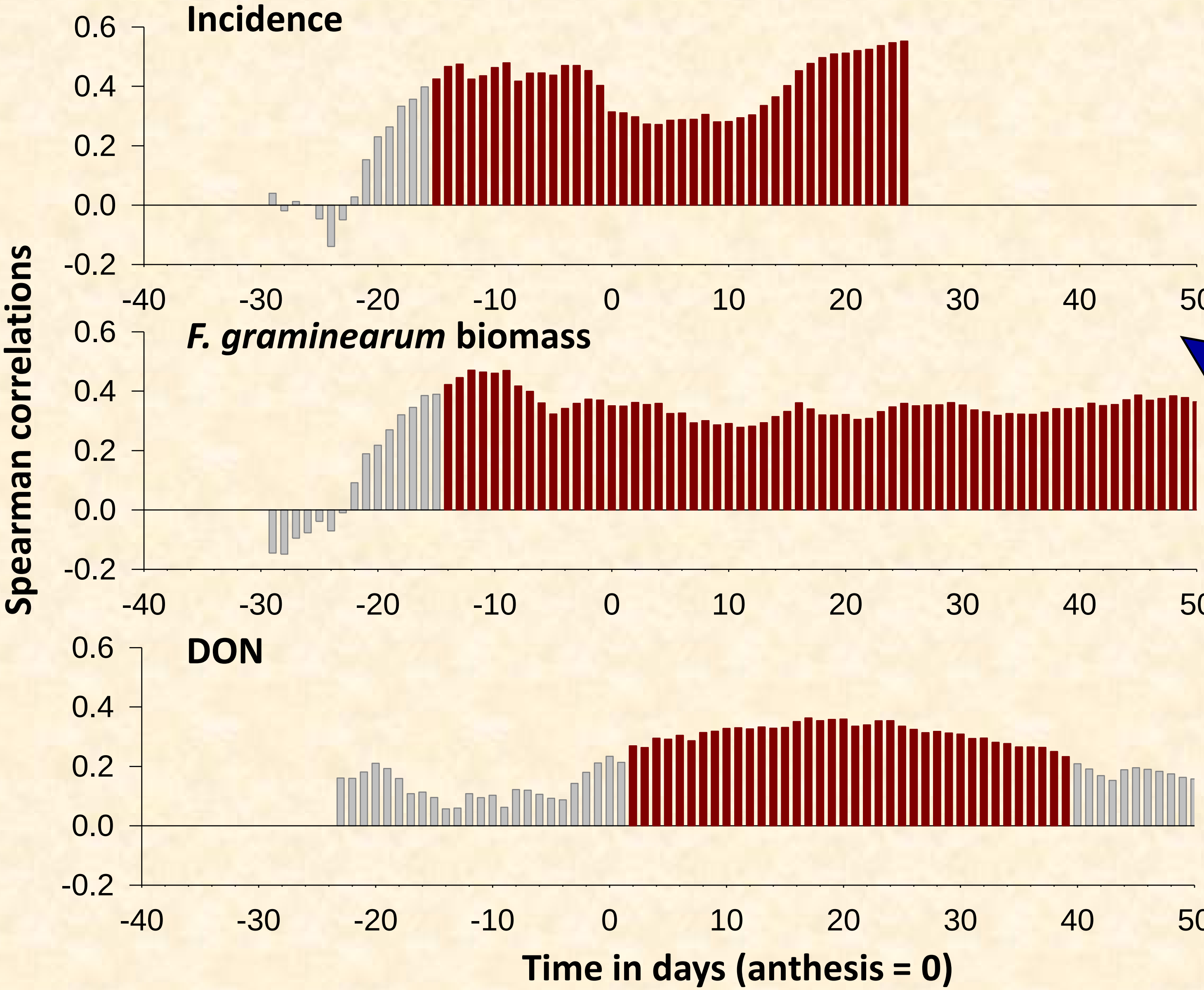


Europe: Average RH and FHB Incidence



Time 0 is now anthesis

Europe: Average RH (15-day windows)



Results not yet available for a wide range of summary environmental variables

Conclusions

- 'Window-Pane' analysis is a useful data-mining tool for exploring the relationships between environment and FHB
  - Although reported results were based on nonparametric correlation coefficients, many test statistics can be utilized
  - It is very important to correct for the multiple-testing problem to avoid *false positives*
  - Numerous specialized (and *ad hoc*) versions of the method are commonly--and successfully--used for analysis, with less formal or less thorough exploration of window length and starting (ending) times
- Many 'summary' environmental variables were correlated with FHB (or with biomass or DON [Europe])
  - (Results for most environmental or response variables & locations were not shown here)
  - As found by others, moisture-wetness-type variables had the highest correlations, for both US (*Kriss et al.*) and European data sets (*unpubl.*)
  - Pure temperature variables had little correlation with FHB for U.S. With European data, moderate negative correlations were found

Conclusions, *continued*

- Relatively short windows during the last 2 months of the season dominated in the results
  - Window-length results did not depend greatly on variable or location
- Results can be used to:
  - Guide the development of the next generation of risk prediction models
  - Separate the direct and indirect effects of environment on DON