The CGMS crop yield forecasting system

Allard de Wit & Raymond van der Wijngaart





About regional crop yield forecasting

- Make an estimate of crop yield or production for the current season before the harvest
- Carried out at the level of administrative regions
- Often calibrated against past regional statistical data
- Continuous "assimilation" of data as the growing season progresses
- Improve accuracy during the growing season
- Better then a baseline forecast (i.e. average or trend)



How do we forecast crop yield





Forecast example: Winter-wheat (Spain)





About crop yield forecasting

"the art of identifying the factors that determine the spatial and inter-annual variability of crop yields" (René Gommes, FAO 2003).



About those factors

"... **AND IF** IT GETS ENOUGH RAIN, **AND** SUN, **AND IF** IT ISN'T KILLED BY HAIL, **AND IF** IT ISN'T DAMAGED BY FROST, **AND IF** WE CAN GET IT OFF BEFORE IT'S COVERED BY SNOW, **AND IF** WE GET IT TO THE ELEVATORS, **AND IF** THE TRAINS ARE RUNNING, **AND IF** THE GRAIN HANDLERS AREN'T IN STRIKE, **AND IF** ... "





Types of forecasting systems

Judgement, based on stakeholders that reach consensus on the expected yield given all available information

- Statistical, based on functional relationships between a crop yield indicator and the crop yield statistics (e.g. time trend models and/or CGMS simulation result)
- Combinations of the above (European MARS system)



Statistical forecasting assumptions

- Uses time-series of historic statistics and crop yield indicators
- Parameterizes a forecasting model explaining the relation using regression.
- The model parameters are derived at several time steps during the growing season (i.e. each dekad)
- Forecasting model is then applied in prognostic mode to forecast the current season's yield



Parameterize the model in time or in space?

- Build a time-series model for one region and multiple years
- Build a spatial model for multiple regions and one year
- Combine the two above (even more difficult!)



Reasons for preferring a time-series model

Several effects:

- Socio-economic factors differ between regions (example: Germany 7.24 ton/ha, Poland 3.44 ton/ha).
- Crop yields often show an upward (or downward) trend over the years
- Simulated year-to-year variability in crop yield differs from variability in regional statistics.



Statistical forecasting models

Parametric models:

- Regression analyses: (multiple -) regression between crop yield statistics and crop indicators
- Non-parametric models:
 - Scenario analyses: Find similar years and use these to forecast
 - Neural networks: train a neural network to recognize yield-indicator relationships



Time-series regression models for crop yield forecasting

- **Basic assumptions:**
- crop yield = f(time-trend + indicators(i1, i2, ...))
- Uses (multiple) linear regression
- Advantages:
- Simple, understandable
- Hypothesis testing (statistical significance)
- Provides models with predictive power



Example of regression analyses for winterwheat in Hebei province



1. time-trend analyses





2. Analyses of residuals





3. Compare deviations with CGMS results

Hebei winter-wheat - Deviations from the trend 0.5 Yield anomaly [ton/ha] 0.4 0.3 0.2 0.1 0 -0.1 -0.2 -0.3 -0.4 -0.5 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 Year





4. Fit linear model through yield anomalies





5. Build and test complete forecasting model

Our complete forecasting model is specified by:

- 1. The time trend model +
- 2. The yield anomaly model as a function of the CGMS simulation results

 $\mathsf{ReportedYield}_y = (a \times y) + b \times CGMS + c$



6. Results of stacked models

Forecast model - Hebei province - winter-wheat





7. Reported vs modelled yields ($R^2 = 0.897$)



Take the time-window too large: unstable trend
Use multiple linear regression with too many indicators (low DF): Good fits but no predictive power



The CGMS statistical toolbox (CST)

Observation:

- Manual analyses is error prone
- Desire by MARS-Stat for a dedicated tool for yield forecasting

Development of CGMS Statistical Toolbox

- CST does several analyses: time trend analyses, (multiple) regression analyses and scenario analyses
- Each model is tested whether it improves prediction beyond the trend only
- Hypothesis testing for determining significance of results



CGMS Statistical Toolbox DEMO

