A Comparison of within season yield predictions algorithms:
On the analysis of the crop model behaviour.

B. Dumont\textsubscript{1}, B. Basso\textsubscript{2}, JP. Destain\textsubscript{1}, B. Bodson\textsubscript{1}, MF. Destain\textsubscript{1}

1, ULg Gembloux Agro-Bio Tech, Gembloux, Belgium
2, Michigan State University & W.K. Kellogg Biological Station, Michigan, USA
I. Introduction
Crop models facts:

- Crop model has been developed for decades now to simulate the crop growth and to assess the impact of cropping systems on the environment and/or vice-versa.

- They have proven to be effective in simulating the water and carbon balances of different/many cropping systems.

- With the advances in spatial information technology, their use has been extended to spatial management of cropping systems, i.e., precision agriculture.

- Crop models are deterministic models: they produce the same output from a given set of starting condition and initial state.

- The decision making process may remains complex (e.g., N management) as it is linked to the non-knowledge of future weather conditions!
In the context of Precision Agriculture, future needs have been highlighted (Mc Bratney et al., 2005):

- “[...] There are not many formal Decision Support Systems (DSS) and no well-designed strategies that are flexible enough to incorporate [...] previously developed practices and concepts into the range of management processes that operate in the practical world.”

- “[...] If we look at the variation of yield across a field and across years, half of the variation comes from year-to-year variation. Knowledge of this temporal aspect needs to be greatly increased. A second issue is within-season management [...] using feedback from crop monitoring is a promising way of optimising inputs.”

In the context of *Precision Agriculture*, future needs have been highlighted (Mc Bratney et al., 2005)

- “ [...] *There are not many formal Decision Support Systems (DSS)* and no well-designed strategies that are flexible enough to incorporate [...] previously developed practices and concepts into the range of management processes that operate in the practical world."

  ➔ To have at one’s disposal a robust methodology to perform in *real-time* the *prediction* of *end-season yields*

- “ [...] If we look at the variation of yield across a field and across years, *half of the variation comes from year-to-year variation*. Knowledge of this temporal aspect needs to be greatly increased. A second issue is *within-season management* [...] using feedback from crop monitoring is a promising way of optimising inputs."

  ➔ To assess the *impacts* on grain *yields* of a wide range of climate conditions representative of the *probable future weather* in a given area

Building a general methodology for real-time prediction of end-season yield:

1 – To simulate the end-season yield, crop models need, as inputs, the climatic data set representative of the whole crop growth season:

2 – Past actual growing conditions can be monitored, updated in real-time, to be used as input of the model

3 – Provided one could find a way to supply the unknown future, the daily updated climatic matrix ensembles could be used to
   • To follow crop growth and
   • To predict end-season yields.
Two validated yield prediction methodologies:

A simple climatic assumption: 

*Seasonal Norm*

Nearly non-limiting conditions 

(Dumont et al., 2013)

Remaining yield potential

A complex climatic assumption: 

*Stochastic approach*

Randomly generated stress events 

(Lawless et al., 2005)

Stochastic yield simulations

---


REAL monitored climate → Past growing conditions and stresses effects
ASSUMED future climate → Potential future growing conditions

Simulations evolves from PURE HYPOTHETIC conditions to PURE REAL conditions

Figure: Model output simulations based on a given Matrix Ensembles 2005-2006 (left) - Evolution of the final output simulation value vs. evolving climatic combination of real and mean climate (right) - Black line (-) : pure hypothetic climate – Grey line ( ) : ME – Dotted black line (--) : pure real climate
II. OBJECTIVES
The goals of this research is:

To determine whether or not the predictive methodologies are equivalent?

To do this, the purpose is to design a generic approach to study model output, which would:

- Rely on a sound theoretical basis
- Highlight a predefined model behaviour, if it exists!
III. MATERIAL & METHODS
The original experiment aims to study the winter wheat growth under different growing conditions:

- Soil types: sandy loam, clay loam and classical loam
- Nitrogen fertilisation level: 0-120-180-240uN
- Nitrogen fertilisation rate: 2 or 3 applications

7 protocols
Material and Method (2/3)

The STICS soil-crop model (Inra, France)

Simulateur multidisciplinaire pour Culture Standard
(Multidisciplinary model for standard culture)
Material and Method (3/3)

Weather database and climate variability

• Original weather database WDB:
  – Site: Ernage (4km from the experimental site)
  – 30 years of data: 1980-2013
  – Royal Meteorological Institute (IRM) of Belgium
Weather database and climate variability

• Original weather database WDB:
  – Site: Ernage (4km from the experimental site)
  – 30 years of data: 1980-2013
  – Royal Meteorological Institute (IRM) of Belgium

• Seasonal norms refer to the value of a climatic variable that occur the most often at a certain time of year (day).
  ➔ It can be easily evaluated, by computing the daily mean values
Material and Method (3/3)

Weather database and climate variability

- Original weather database WDB:
  - Site: Ernage (4km from the experimental site)
  - 30 years of data: 1980-2013
  - Royal Meteorological Institute (IRM) of Belgium

- Seasonal norms refer to the value of a climatic variable that occur the most often at a certain time of year (day).

- Stochastic weather generator: LARS-WG
  - Generate synthetic climates according to original database characteristics

30-years WDB → 300 Synthetic-Stochastic climates
The Convergence in Law Theorem:

Formulation 1:
Let \( \{X_n\} \) be a sequence of \( n \) random variables and let \( X \) be a random variable. Denote by \( F_n(x) \) the distribution function of \( X_n \).

Then \( \{X_n\} \) converges in distribution to \( X \) as \( n \to \infty \) (\( X_n \to_L X \) or \( X_n \to_d X \)), if and only if there exists a distribution function \( F_X(x) \) such that the sequence \( \{F_n(x)\} \) converges to \( F_X(x) \) for all points \( x \in \mathbb{R} \) where \( F_X(x) \) is continuous.

Formulation 2:

\[
X_n \to_L X \text{ as } n \to \infty,
\]
if, there exists a function \( f: \mathbb{R} \to \mathbb{R} \), continuous and bounded such that:

\[
E[f(X_n)] \xrightarrow{\mathcal{L}} E[f(X)]
\]
The Central Limit Theorem (CLT)

Let \( \{Y_n\} \) be independent random variables, of same law, of integrable square. We denote \( \mu \) its expectation and \( \sigma^2 \) its finite variance (here we assume that \( \sigma^2 > 0 \)).

Then

\[
\frac{\sqrt{n}}{\sigma} \left( \frac{\sum_{i=1}^{n} Y_n}{n} - \mu \right) \xrightarrow{L} Y, \quad \text{when } n \to \infty
\]

In practical terms, the CLT implies that for “large” \( n \), the distribution of \( Y_n \) may be approximated by a Gaussian distribution with mean \( \mu \) and variance \( \sigma^2/n \).
The Generalised Central Limit Theorem (GCLT)

1 - Under usual agronomic practices, crop cereal yields have rarely been found to respond to a normal distribution

2 - Even under the most favourable climatic conditions, such crop will exhibit natural genetic limitations to yield elaboration

3 - Richard Day (1965) suggested to apply a log-normal transformation to the study of yields distributions.

\[ Y_n \sim \log \left( \max_{n} Y \right) = \ln Y_{\max} - Y_n \sim Y_{\max} \]

4 - The CLT admits different generalizations

➤ The log. of a product equals the sum of the log. of its factors.

A process based on sound statistical foundations
Let's consider the following schema block based on the Convergence in Law Theorem

\[
X_n \xrightarrow{L} X \text{ as } n \to \infty
\]

- \(X_n \to\) Climatic inputs (#300) stochastically generated according to Lawless et al. (2005)
- \(X\) \to Mean projected climatic data used by Dumont et al. (2013)
A process based on sound statistical foundations
Let's consider the following schema block based on the Convergence in Law Theorem

\[ X_n \rightarrow_L X \text{ as } n \rightarrow \infty \]

\[ \xrightarrow{\text{LARS-WG \& CLT}} \]

\[ X_n \rightarrow \text{Climatic inputs (#300) stochastically generated according to Lawless et al. (2005)} \]

\[ X \rightarrow \text{Mean projected climatic data used by Dumont et al. (2013)} \]
A process based on sound statistical foundations

Let's consider the following schema block based on the Convergence in Law Theorem:

\[ X_n \overset{L}{\to} X \text{ as } n \to \infty \]

\[ E \left[ \sum_{n} \right] \to E \left[ \right] \]

\[ Y(t+\Delta t) = Y(t) + f(Y(t), X(t), \theta) \]

\[-1\] LARS-WG & CLT

\[-2\] STICS behaviour & GCLT

\[ X_n \to \text{Climatic inputs (#300) stochastically generated according to Lawless et al. (2005)} \]

\[ X \to \text{Mean projected climatic data used by Dumont et al. (2013)} \]
A process based on sound statistical foundations

Let’s consider the following schema block based on the Convergence in Law Theorem

\[ X_n \xrightarrow{L} X \text{ as } n \to \infty \]

- 1 - LARS-WG & CLT
- 3 - Yield distribution & Predictive ability

\[ Y(t + \Delta t) = Y(t) + f(Y(t), X(t), \theta) \]

\[ X_n \to \text{Climatic inputs (#300) stochastically generated according to Lawless et al. (2005)} \]

\[ X \to \text{Mean projected climatic data used by Dumont et al. (2013)} \]
A process based on sound statistical foundations

Let's consider the following schema block based on the Convergence in Law Theorem

\[ X_n \rightarrow_L X \text{ as } n \rightarrow \infty \]

- 1 - LARS-WG & CLT
- 3 - Yield distribution & Predictive ability

\[ E \left( \sum_{i=1}^{n} X_i \right) \rightarrow E \left( \sum_{i=1}^{n} f(Y(t_i), X(t_i), \theta) \right) \]

- 2 - STICS behaviour & GCLT

\[ Y(t+\Delta t) = Y(t) + f(Y(t), X(t), \theta) \]

\[ X_n \rightarrow \text{Climatic inputs (#300) stochastically generated according to Lawless et al. (2005)} \]

\[ X \rightarrow \text{Mean projected climatic data used by Dumont et al. (2013)} \]
V. RESULTS
The LARS-WG & Central Limit Theorem

"The generator option of LARS-WG may be used to generate synthetic data which have the same statistical characteristics as the observed weather data" (Semenov and Barrow, 1997, 2002).

\[
X_n \xrightarrow{L} X \quad \text{as} \quad n \to \infty
\]

Yield distribution & GCLT

\[
Y(t + \Delta t) = Y(t) + f(Y(t), X(t), \theta)
\]

RESULTS (2/8)

**STICS behaviour & GCLT**

Let’s consider one have no information about the future:

\[ Y_{\text{max}} = \text{Yield max. observed under the 300 climates (14.5 ton.ha}^{-1}) \]

KS-test: \( p\text{-value} = 0.837 > 0.025 (= \alpha/2) \)

\[ y \Rightarrow \text{The experimental distribution CAN NOT be consider as significantly different from a log-normal distribution} \]
Yield distribution & GCLT

STICS crop model could be considered as a global \textit{f-function} (log-normal) that links together the $X(t)$ random climatic input and the $Y(t)$ simulated grain yield output.

$$Y(t+\Delta t) = Y(t) + f(Y(t), X(t), \theta)$$

\begin{align*}
LARS-WG & \text{ & CLT} \\
\text{Yield distribution & Predictive ability}
\end{align*}
Results (4/8)

Yield distribution & Predictive ability

Let's consider the predictive methodology of L&S:

Extended KS-test to different combination of measured and projected climates

Season 1981-82
Yield distribution & Predictive ability

Let's consider the predictive methodology of L&S:

- For the main part of the season, in a predictive approach, the STICS crop model could be considered as a global \( f \)-function that links together the \( X(t) \) random climatic input and the \( Y(t) \) simulated grain yield output.

\[
E \xrightarrow{\theta} E \xrightarrow{\Delta t} Y(t + \Delta t) = Y(t) + f(Y(t), X(t), \Theta)
\]
Results (6/8)

Yield distribution & Predictive ability

Comparison of the simulations using the two projective assumptions:

- Mean projective climate assumption (Seasonnal norms)
- Mean of the simulations using stochastic approach
--- 95% confidence interval
Results (7/8)

Yield distribution & Predictive ability

Comparison of the simulations using both projective assumptions:

- 90% of the time the simulated grain yield were judged equivalent with $RRMSE < 10\%$ & $ND < 10\%$ between methods.

<table>
<thead>
<tr>
<th>Year</th>
<th>RRMSE</th>
<th>ND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.044</td>
<td>-0.034</td>
</tr>
<tr>
<td>2</td>
<td>0.038</td>
<td>-0.022</td>
</tr>
<tr>
<td>3</td>
<td>0.047</td>
<td>-0.017</td>
</tr>
<tr>
<td>4</td>
<td>0.057</td>
<td>-0.007</td>
</tr>
<tr>
<td>5</td>
<td>0.061</td>
<td>-0.007</td>
</tr>
<tr>
<td>6</td>
<td>0.062</td>
<td>-0.006</td>
</tr>
<tr>
<td>7</td>
<td>0.075</td>
<td>-0.006</td>
</tr>
<tr>
<td>8</td>
<td>0.079</td>
<td>-0.005</td>
</tr>
<tr>
<td>9</td>
<td>0.081</td>
<td>-0.005</td>
</tr>
<tr>
<td>10</td>
<td>0.086</td>
<td>-0.004</td>
</tr>
<tr>
<td>11</td>
<td>0.095</td>
<td>-0.003</td>
</tr>
<tr>
<td>12</td>
<td>0.102</td>
<td>-0.002</td>
</tr>
<tr>
<td>13</td>
<td>0.103</td>
<td>-0.002</td>
</tr>
<tr>
<td>14</td>
<td>0.109</td>
<td>-0.001</td>
</tr>
<tr>
<td>15</td>
<td>0.111</td>
<td>-0.001</td>
</tr>
<tr>
<td>16</td>
<td>0.114</td>
<td>-0.001</td>
</tr>
<tr>
<td>17</td>
<td>0.116</td>
<td>-0.001</td>
</tr>
<tr>
<td>18</td>
<td>0.117</td>
<td>-0.001</td>
</tr>
<tr>
<td>19</td>
<td>0.118</td>
<td>-0.001</td>
</tr>
<tr>
<td>20</td>
<td>0.119</td>
<td>-0.001</td>
</tr>
<tr>
<td>21</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>22</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>23</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>24</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>25</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>26</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>27</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>28</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>29</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>30</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>31</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>32</td>
<td>0.120</td>
<td>-0.001</td>
</tr>
<tr>
<td>Global</td>
<td>0.1124</td>
<td>-0.058</td>
</tr>
</tbody>
</table>

$Grain \ Yield \ [\text{ton} \cdot \text{ha}^{-1}]$ obtained using Demirel et al. (2012) approach.

$Grain \ Yield \ [\text{ton} \cdot \text{ha}^{-1}]$ obtained using mean value of Lawoko et al. (2005) approach.
Results (8/8)

Yield distribution & Predictive ability

→ *Lead-time predictions* were found *equivalent* for significant level of yield prediction.

\[
X_n \xrightarrow{L} X \text{ as } n \to \infty
\]

Yield distribution & CLT

\[
Y(t + \Delta t) = Y(t) + f(Y(t), X(t), \theta)
\]

LARS-WG & CLT

Yield distribution & Predictive ability

\[
E \left[ X_n \right] \to E \left[ X \right]
\]
VI. Conclusions
Conclusions

A demonstration? Not properly!

However for the considered case study

- A wheat crop culture
- Sown on classic loamy soil
- Growing under a temperate climate

The proposed methodology allowed to conclude that

- The convergence in law theorem was of application
- The central limit theorem was of application

⇒ The two predictive approaches are equivalent

⇒ Higher efficiency of Mean Climate assumption
- Using seasonal norms as projection is 300 times faster -

!!! Computational time is important in DSS !!!
A Comparison of within season yield predictions algorithms:  
On the analysis of the crop model behaviour.

Benjamin Dumont  
Research Assistant & PhD Student @ GxABT-ULg  
Benjamin.dumont@ulg.ac.be