Using Automatic Differentiation to study the sensitivity of a crop model

Lauvernet, C.    Hascoët, L.    Le Dimet, F.-X.    Baret, F.

Irstea (Environment and Agriculture)
INRIA (Computer Science and Control)
INRA (Agronomy)

6th International Conference on Automatic Differentiation
Contents

1 The application domain: the agronomic crop model STICS

2 Sensitivity analysis of STICS
   - Sensitivity analysis: what it is and what for?
   - Selected input parameters for sensitivity analysis

3 Automatic Differentiation of STICS
   - The TAPENADE Automatic Differentiation tool
   - STICS adjoint: the pains and sufferings of an AD end-user
   - Validation of the adjoint model

4 Conclusion
   - Sensitivity results of LAI and biomass
   - Benefits
Plan

1. The application domain: the agronomic crop model STICS
2. Sensitivity analysis of STICS
   - Sensitivity analysis: what it is and what for?
   - Selected input parameters for sensitivity analysis
3. Automatic Differentiation of STICS
   - The TAPENADE Automatic Differentiation tool
   - STICS adjoint: the pains and sufferings of an AD end-user
   - Validation of the adjoint model
4. Conclusion
   - Sensitivity results of LAI and biomass
   - Benefits
**The application domain: the agronomic crop model STICS**

STICS simulates the behaviour of the soil-crop system over one or successive crop cycles.

---

**Objectives**

to simulate consequences on crop production and environment of variations in: climate, soil, and crop management

**Simulated object (Control Parameters)**
a cultural situation = a soil-crop system + a technical itinerary

**Simulated processes (Outputs)**

⇝ growth and development of the crop
⇝ water and nitrogen balance of the soil-crop system
Key output variables of STICS

**LAI: Leaf Area Index**

Total one-sided area of leaf tissue per area of ground surface.

Leaves are the main interface with the atmosphere for the transfer of mass and energy \( \Rightarrow \) the LAI indirectly describes:

- potential of photosynthesis available for primary production
- plant respiration, evapotranspiration
- biosphere ↔ atmosphere carbon flux.
- severely affected areas (fires, parasites...)

LAI closely related to:

**Biomass**

Total mass of living matter per area of land.
Plan

1. The application domain: the agronomic crop model STICS

2. Sensitivity analysis of STICS
   - Sensitivity analysis: what it is and what for?
   - Selected input parameters for sensitivity analysis

3. Automatic Differentiation of STICS
   - The TAPENADE Automatic Differentiation tool
   - STICS adjoint: the pains and sufferings of an AD end-user
   - Validation of the adjoint model

4. Conclusion
   - Sensitivity results of LAI and biomass
   - Benefits
Sensitivity analysis is an initial goal

⇝ further goals include (≈200) parameter estimation

Sensitivity analysis studies the impact of perturbing the control (input) parameters on the model output. It helps to:

- prioritize the model parameters
- identify critical regions in the space of the inputs (including interactions)
- support decision for research and future experiments.
  ⇝ which parameters deserve accurate measurements?
- gauge model adequacy and physical relevance
- simplify models (metamodelling)
- find technical errors in the model
Sensitivity analysis: mathematical framework

- A model:
  \[ F(X, K) = 0 \]

  \( X \) = state variables (LAI, biomass . . . )
  \( K \) = control variables (parameters, forcing variables . . . )

  Here, the model is exactly the STICS computer program \( K \mapsto X \).

- A response function \( G(X) \), e.g.
  \[ G_{\text{LAI}} = \sum_{i=1}^{T} \text{LAI}(t_i) \]
  or
  \[ G_{\text{biomass}} = \sum_{i=1}^{T} \text{biomass}(t_i) \]

The problem is to evaluate the sensitivity of \( G \) with respect to \( K \)
or in other words the gradient of \( G \) with respect to \( K \).

With an adjoint model, only 2 steps:
1. run the direct model once for the given \( K \)
2. solve the adjoint model

\[ \nabla G = \frac{dG}{dK}^t = \left( \frac{dG}{dX} \cdot \frac{dX}{dK} \right)^t = \left( \frac{dX}{dK} \right)^t \cdot \left( \frac{dG}{dX} \right)^t \]
Sensitivity analysis: adjoint model or other approaches?

Adjoint method is the only way to calculate formally the gradient of the response function at a cost that does not depend on the size of $K \Rightarrow$ appropriate when number of entries $K \gg$ size of response function.

<table>
<thead>
<tr>
<th>other approach</th>
<th>advantages</th>
<th>disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>tangent-linear</td>
<td>no need to calculate the adjoint</td>
<td>cost proportional to the size of $K$</td>
</tr>
<tr>
<td>finite difference</td>
<td>easy calculations</td>
<td>approximation of the gradient + extensive direct model computations</td>
</tr>
<tr>
<td>stochastic sampling</td>
<td>global sensitivity</td>
<td>cost grows rapidly with the dimension of $K$</td>
</tr>
</tbody>
</table>
Practical problems with models like STICS

Piecewise differentiable function
⇒ only left- and right-derivatives (or sub-gradients)

〜 Derivative-free methods (divided differences, stochastic, ...) behave better in these cases, but at a cost.
〜 In practice, problem is overlooked: local sensitivity valid in a neighborhood of $K$. 

Lauvernet, C., Hascoët, L. et al. Automatic Differentiation for crop modeling
### Selected input parameters for sensitivity analysis

<table>
<thead>
<tr>
<th>parameter</th>
<th>definition</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dlaimaxbrut</td>
<td>maximum rate of gross leaf surface area production</td>
<td>0.00044</td>
</tr>
<tr>
<td>stlevamf</td>
<td>cumulated development units between the LEV and AMF stages</td>
<td>208.298</td>
</tr>
<tr>
<td>stamflax</td>
<td>cumulated development units between AMF and LAX stages</td>
<td>181.688</td>
</tr>
<tr>
<td>jvc</td>
<td>days of vernalisation (cold days needed to lift)</td>
<td>35</td>
</tr>
<tr>
<td>durvieF</td>
<td>lifespan of a cm of adult leaf</td>
<td>160</td>
</tr>
<tr>
<td>adens</td>
<td>compensation between number of stems and plants density</td>
<td>-0.6</td>
</tr>
<tr>
<td>efcroijuv</td>
<td>maximum growth efficiency during juvenile phase (LEV-AMF)</td>
<td>2.2</td>
</tr>
<tr>
<td>efcroiveg</td>
<td>maximum growth efficiency during vegetative phase (AMF-DRP)</td>
<td>2.2</td>
</tr>
<tr>
<td>efcroirepro</td>
<td>maximum growth efficiency during grain filling phase (DRP-MAT)</td>
<td>4.25</td>
</tr>
<tr>
<td>vmax2</td>
<td>maximum rate of nitrate absorption by the roots</td>
<td>0.05</td>
</tr>
</tbody>
</table>

All these parameters would deserve a good parameter estimation...
Plan

1. The application domain: the agronomic crop model STICS

2. Sensitivity analysis of STICS
   - Sensitivity analysis: what it is and what for?
   - Selected input parameters for sensitivity analysis

3. Automatic Differentiation of STICS
   - The TAPENADE Automatic Differentiation tool
   - STICS adjoint: the pains and sufferings of an AD end-user
   - Validation of the adjoint model

4. Conclusion
   - Sensitivity results of LAI and biomass
   - Benefits
We build the adjoint of STICS with TAPENADE.

TAPENADE (2.0 at the time, now 3.6):
- Automatic Differentiation by source-to-source transformation
- Flow- and Context-sensitive global data-flow analysis
- Association by name
- Adjoint with Store-All strategy
- Checkpointing on calls by default
A few recommended programming practices

STICS is FORTRAN 77 ⇒ In theory, TAPENADE can build its adjoint
  ⇔ but improvements required on old TAPENADE (< 2005),
  ⇔ plus some adaption of STICS

Adaption of STICS source:
- Active vars should be separate from others: no big work array!
- Active independents and dependents must be clearly identified
  ⇒ keep them separate from the other variables.
- Use a coherent precision level: DOUBLE PRECISION
- Solve portability problems unveiled by AD:
  no uninitialized remanent globals!
Issues due to AD in general

- *the program itself is the equation* (no ODE/PDE)
  ⇒ conditional jumps with no underlying continuity/differentiability.
- Uncontrolled code evolution
  ⇒ fast increase of new subcases and subdivisions.

Several subroutines needed manual improvement to Restore differentiability
⇒ a cleaner STICS?
Issues due to TAPENADE

- Remarkably large number of branches: thresholds, conditions, loops... that must be stored for control-flow reversal.
  ⇒ define PUSHCONTROL(), cheaper than storing a full INTEGER

- Time stepping: split main time loop (400 steps) into nested 20-steps loops. Make them subroutines to force checkpointing
  ⇒ now with TAPENADE 3.6, use $AD BINOMIAL-CKP

- Fine-tune checkpointing of nested calls:

⇒ $AD NOCHECKPOINT

Co-evolution of STICS and TAPENADE
⇒ a more portable STICS & a more efficient adjoint code
Validation of the adjoint model

- **Divided Differences:** 0.42248278309969720000
- **Tangent AD:** 0.42248278406221984000
- **Adjoint AD:** 0.42248278406222007000

- Performance with TAPENADE 2.0 (2005):
  - Orig: 0.21 s  Tgt: 0.39 s  Adj: 30.96 s  Traffic: 13.8 Gb  Peak: 240 Mb

- After checkpointing fine tuning (**C$AD  NOCHECKPOINT**)
  - Orig: 0.22 s  Tgt: 0.52 s  Adj: 0.86 s  Traffic: 0.2 Gb  Peak: 162 Mb
Plan

1. The application domain: the agronomic crop model STICS
2. Sensitivity analysis of STICS
   - Sensitivity analysis: what it is and what for?
   - Selected input parameters for sensitivity analysis
3. Automatic Differentiation of STICS
   - The TAPENADE Automatic Differentiation tool
   - STICS adjoint: the pains and sufferings of an AD end-user
   - Validation of the adjoint model
4. Conclusion
   - Sensitivity results of LAI and biomass
   - Benefits
Hierarchy of influential parameters:

- **LAI**: *adens* is the most influential parameter (47%). It represents the ability of a plant to withstand increasing densities, depends on the species and varieties ⇒ *strong influence* for this type of wheat and less for other crops.

- **Biomass**: hierarchy modified by the strong influence of the efficiency *efcroiveg* (maximum growth efficiency during vegetative phase) ⇒ *we can ignore the estimate of* *efcroiveg* if we only work on LAI, but absolutely not to simulate biomass.

- relatively low sensitivity (5% and 3%) of biomass integrated over the life cycle to the other 2 parameters of efficiency *efcroirepro* and *efcroijuv* ⇒ *the biomass is not so dependant on the juvenile and the grain filling phases but essentially on the vegetative phase.*

- **LAI** is actually dependant on 4 params and biomass on 5 ⇒ *the user should concentrate on these and estimate them better.* Uncertainty on the other parameters is of smaller importance.
Conclusion

- For the agronomic community, the adjoint model of STICS is an efficient way to perform sensitivity analysis since it requires the calculation only once for each agro-pedo-climatic situation.
- AD is a choice approach for sensitivity analysis of agronomic models, where the code is the model.
- Sensitivity analysis focuses user’s attention on some parameters and modules, according to the user’s objectives.
- ! This local sensitivity analysis is local, valid only in a small neighborhood and the hierarchy of sensitivities may vary under different conditions (but in practice: quite stable hierarchy).
- Sensitivity analysis is a preliminary to parameter estimation, data assimilation: many of these agronomic parameters (yield, balance,...) are not directly observable.
- This shows that variational methods are applicable in agronomy but it takes a truly pluridisciplinary work!