SUMO WP2
Learning low complexity and intermediate complexity models

Wim Wiegerinck
Annual review meeting
Skopje, November 24, 2011
Objectives WP2

• Research and develop efficient, robust and scalable learning strategies to optimize connection coefficients for dynamical systems of low and intermediate size complexity (up to 1000 variables)

• Research estimators for model performance (later).
Research methods in WP2

• Mainly focused on
  – Exploring and improving existing learning strategies.
  – Methods will be experimentally assessed on
    • performance on systems agreed upon in WP1
    • their scalability to high dimensional climate models

• Later in the project
  – Feedback from WP3-5 will further guide the research directions in this WP.
  – Issues that arise in those work packages will be referred back to WP1 and this WP for further analysis and improvement.
Task period 1

Study and comparison of different SUMO learning algorithms:

• Design and research methods to optimize the connection coefficients in the supermodel.

• Assessment on
  – Performance: plots & the metrics defined in WP1
  – Computational scalability to systems with many degrees of freedom.
Two subtasks

Subtask 2.1. learning in ODEs - simplest case
• All variables observed
• Similar time scales

Subtask 2.2. learning in ODEs - more complex
• Incomplete data, incomplete knowledge
• Constraints on the connections
• Time varying parameters
• Slow and fast time scales
Subtask 2.1

• Models
• Procedure
• Supermodels (SUMO)
• Learning methods
• Results
  – Lot of pictures
  – Training times
• Conclusions
Models

- Lorenz 63 (3 dof)
- Rössler (3 dof)
- Lorenz 84 (3 dof)
- low dimensional barotropic channel model (6 dof)
- T5 quasi geostrophic model (30 dof)
Procedure

• ODE: $\dot{x} = f(x, \theta)$
• Ground truth: parameters $\theta_{gt}$
• Imperfect models: parameters $\theta_{\mu}$
• SUMO ODE: $\dot{x}_{\text{sumo}} = F(x_{\text{sumo}}, W, \{f(., \theta_{\mu})\})$
• Optimize $W$ on training set
  — record required time
• Assess performance SUMO ODE on test set
  — *Attractor* metrics (not short term prediction)
SUMO types

Connected SUMO:

\[ \dot{x}^i_{\mu} = f^i_{\mu}(x_{\mu}) + \sum_{\nu} C^i_{\mu\nu}(x^i_{\nu} - x^i_{\mu}), \quad C^i_{\mu\nu} \geq 0 \]

\[ \overline{x}_{\text{sumo}} = \frac{1}{M} \sum_{\mu} x_{\mu} \]

Weighted SUMO:

\[ \dot{x}^i = \sum_{\mu} w^i_{\mu} f^i_{\mu}(x), \quad w^i_{\mu} \geq 0, \quad \sum_{\mu} w^i_{\mu} = 1 \]
SUMO Learning: nudging

Nudging – coupling to ground truth

\[ \dot{x}_\mu^i = f_\mu^i(x_\mu) + \sum_{\nu \neq \mu} C_{\mu\nu}^i (x_\nu^i - x_\mu^i) + K (x_{gt}^i(t) - x_\mu^i) \]

\[ \dot{C}_{\mu\nu}^i = a (x_\nu^i - x_\mu^i) \left( x_{gt}^i(t) - \frac{1}{M} \sum_\mu x_\mu^i \right) - \epsilon / (C_{\mu\nu}^i - C_{\text{max}})^2 + \epsilon / (C_{\mu\nu}^i + \delta)^2 \]

Run iteratively until convergence
SUMO learning: cost function

\[ E(W_{\text{sumo}}) = \frac{1}{K\Delta} \sum_{i=1}^{K} \int_{t_i}^{t_i+\Delta} (\bar{x}_{\text{sumo}}(W_{\text{sumo}}, t) - x_{gt}(t))^2 dt \]

- Applicable to both SUMOs
- Use your favorite minimizer
- Iterative
SUMO one-shot learning

Error of imperfect model $\mu$ for each $t$

$$\epsilon^i_\mu(t) = f^i_\mu(x^t_{gt}(t))dt - (x^i_{gt}(t + dt) - x^i_{gt}(t))$$

Error of weighted SUMO for each $t$

$$\epsilon^i(t) = \sum_\mu w^i_\mu \epsilon^i_\mu(t)$$

Total squared error of weighted SUMO

$$E^i(w^i) = \sum_t (\epsilon^i(t))^2 = \sum_{\mu\nu} w^i_\mu w^i_\nu (\sum_t \epsilon^i_\mu(t)\epsilon^i_\nu(t))$$

$$= \sum_{\mu\nu} w^i_\mu w^i_\nu \chi_{\mu\nu}$$

- Quadratic programming
- No repeated run of ODEs
Lorenz 63 imperfect models

Attractor plots:
• Blue: ground truth
• Red: imperfect models
Lorenz 63 SUMO

Ensemble mean
Mean dynamics
Connected+nudging

Connected+cost
Weighted+cost
Weighted+oneshot
Rössler imperfect models

Attractor plots:
• Blue: ground truth
• Red: imperfect models
Rössler SUMO

Ensemble mean
Mean dynamics
Connected+nudging

Connected+cost
Weighted+cost
Weighted+oneshot
Lorenz 84 imperfect models

Attractor plots:
- Blue: ground truth
- Red: imperfect models
Lorenz 84 SUMO

Ensemble mean
Mean dynamics
Connected+nudging
Connected+cost
Weighted+cost
Weighted+oneshot
6 dof barotropic imperfect models

Attractor plots: 2-d projection
- Blue: ground truth
- Red: imperfect models
6 dof barotropic SUMO

Ensemble mean

Mean dynamics

Connected+nudging

Connected+cost

Weighted+cost

Weighted+oneshot
Low order barotropic SUMO time series

ground truth

C optimized with cost function

w, one-shot optimized
T5 QG  imperfect models

Attractor plots: 2-d projection
• Blue: ground truth
• Red: imperfect models
T5 QG SUMO

SUMO, equal weights

SUMO, optimized weights
SUMO metrics

Low dimensional systems:
• KL metric ~ pictures

High dimensional systems:
• Pictures: only from projections
• KL metric quantifies attractor distance
### SUMO training times

<table>
<thead>
<tr>
<th></th>
<th>Lnz 63</th>
<th>Rossler</th>
<th>Lnz 84</th>
<th>LO Btr</th>
</tr>
</thead>
<tbody>
<tr>
<td>C nudge</td>
<td>86</td>
<td>43</td>
<td>126</td>
<td>-</td>
</tr>
<tr>
<td>C cost</td>
<td>590</td>
<td>450</td>
<td>104</td>
<td>531</td>
</tr>
<tr>
<td>W cost</td>
<td>12</td>
<td>17</td>
<td>17</td>
<td>619</td>
</tr>
<tr>
<td>W 1-sht</td>
<td>1.6</td>
<td>1.6</td>
<td>2.1</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Lorenz 63 with double $\rho$

What happens if you change $\rho$ in ground truth and imperfect models in SUMO?
• No retraining of SUMO!

• Result: SUMO's response follows the ground truth
Conclusions task 2.1

- All methods provide in general good quality SUMOs
- Best method depends on problem
  - Weighted SUMOs are less complex
  - Connected SUMOs are more flexible
- SUMO response to parameter change
- Scalability: one-shot method
  - No iterations of ODE integrations required
  - Optimization = low dimensional QP
  - *Can work in real models!*
SUMO relations

• Connected SUMOs \( \leftrightarrow \) weighted SUMOs

• Explanation SUMOs response to parameter change

• And more...
Subtask 2.2

- Parametrized imperfect models
  - Imperfect models in different model class
  - Missing variables
  - SUMO as previous → but partially coupled
  - Lorenz 63

- Limited information exchange
  - OASIS
  - Coarse grained training data
  - New SUMO
  - Lorenz 63
Parametrized imperfect models

- Not all variables explicitly modelled
- Parametrizations for missing variables

\[
\begin{align*}
\dot{x}_a &= \sigma(y_a - x_a) \\
\dot{y}_a &= x_a(\rho - Z(x_a, y_a)) - y_a \\
\dot{y}_b &= X(y_b, z_b)(\rho - z_b) - y_b \\
\dot{z}_b &= X(y_b, z_b)y_b - \beta z_b 
\end{align*}
\]

$Z$ and $X$ : functions, e.g. learned from data
Vector fields and attractors

Blue: Model a  
(x,y)

Red: GT  
(x,y) proj.

Blue: Model b  
(y,z)

Red: GT  
(y,z) proj.
SUMOs

• Connected and weighted (only in $y$)
SUMOs

Connected

Weighted
Conclusions

• Connected & weighted SUMO still works with more difficult parametrized imperfect models
• Partly coupled
• Connected models seem to perform a bit better (in this example). More in the afternoon
• Will be explored more in next period.
Limited information exchange

• Ensemble of imperfect models.
• Reset at weighted average at discrete times
  – No ODE

• Learning similar to one-shot learning

• Hardly degrades (L 63)
Highlights of the results

• SUMO works in low dof ODEs
• Novel SUMO with scalable(!) learning
  – may be feasible even for real climate models
• Oasis-like coupling can be applied
• New `modeling’ of imperfect models
What is next?

• T 2.1 : Upscaling of SUMO $\rightarrow O(1000)$ or more variables

• T 2.2.
  – Explore novel imperfect model modeling
  – Partly coupled models
  – OASIS-like setting
  – Models with different time scales
  – Time varying models
  – SUMO learning under more parameter constraints
  – Uncertainty