HOW TO CHOOSE THE BEST SIMULATION FOR A SPECIFIC PURPOSE?

A. Martilli, J. L. Santiago (CIEMAT, Spain)
T. G. Reisin, (SOREQ, Israel)
A. Baklanov, R. Nuterman (DMI, Denmark)
J. Bartzis, G. Efthimiou (U. W. Macedonia, Greece)
R. Buccolieri, and S. Di Sabatino (U. Salento, Italy)
A. M. Costa, R. Tavares (U. Aveiro, Portugal)
J. Franke, (U. Siegen, Germany)
A. Hellsten, FMI, Finland

Contact: alberto.martilli@ciemat.es
Why do we use models?
Why do we use models?

Because we want to know the value of quantities of interest (QI) for which there is no experimental information.
Why do we use models?

Because we want to know the value of quantities of interest (QI) for which there is no experimental information.

Examples: 3D distribution of pollutants, area with concentrations above a threshold, maximum concentrations, impact of a reduction strategy, tomorrow’s air quality, etc.

QI depends on the purpose of model use.
In modelling activity there are many uncertainties. (parameterizations, numerics, initial and boundary conditions, etc.)
In modelling activity there are many uncertainties. (parameterizations, numerics, initial and boundary conditions, etc.)

For the same case study an ensemble of simulations can be produced.
In modelling activity there are many uncertainties. (parameterizations, numerics, initial and boundary conditions, etc.)

For the same case study an ensemble of simulations can be produced.

How to select those simulations that are fit for purpose?
This decision must be based on a measure of the “distance” between the real world value of $QI$ and the simulated value $(SQI)$.

$$d_{\text{purpose}}(QI, SQI)$$

...and a quantitative criterion of acceptance ($H$) based on the purpose.
This decision must be based on a measure of the “distance” between the real world value of $QI$ and the simulated value $(SQI)$. $d_{purpose}(QI, SQI)$

...and a quantitative criterium of acceptance $(H)$ based on the purpose.

Ex. $QI=C$ (averaged concentration in a certain area). 

$$d_{purpose} = 100 \left( \frac{|C - SC|}{(C + SC)/2} \right) , \quad H=50\%$$
But.. the problem is that $d_{\text{purpose}} (QI, SQI)$ cannot be computed because there is no experimental information on $QI$. 
But.. the problem is that $d_{\text{purpose}} (QI, SQI)$ cannot be computed because there is no experimental information on $QI$.

What can be computed are distances between experimental quantities and the correspondent simulated values $d_X(EQ, SEQ)$.
But.. the problem is that $d_{purpose} (QI, SQI)$ cannot be computed because there is no experimental information on $QI$.

What can be computed are distances between experimental quantities and the correspondent simulated values $d_X(EQ, SEQ)$

What's the value of $K$???

E. g. standard metrics (RMSE, Hit Rate, Fractional Bias, Factor of 2, etc.) or new ones can be created

Ex. EQ=A (averaged concentration at the measurement points).

$$d_x = 100 \frac{|A - SA|}{(A + SA) / 2}$$

or $d_X=RMSE, Fractional~Bias,~etc.$
The problem is to select the best metrics $d_X$ that can surrogate $d_{\text{purpose}}$.

We want a $d_X$ such that:

$$d_X(SEQ_i, EQ) > d_X(SEQ_j, EQ) \iff d_{\text{purpose}}(SQI_i, QI) > d_{\text{purpose}}(SQI_j, QI)$$

We want a separation value $K$ such that:

$$d_X(SEQ_i, EQ) < K \iff d_{\text{purpose}}(SQI_i, QI) < H$$

How to compare metrics?
Use the ensemble of simulations to compare metrics. For each couple of simulations \((i,j)\), estimate:

\[
d_{\text{purpose}}(SQI_i, SQI_j)
\]

\[
d_X(SEQ_i, SEQ_j)
\]

And base the comparison between metrics on the following two techniques:

- Models must have passed a scientific evaluation
1) Kendall’s TAU – it measures the similarity between rankings

\[ \tau = \frac{n_t - n_f}{N^4} \]

The highest the value of \( \tau \), the most similar are the rankings

\[
\begin{align*}
\{ & d_X(SEQ_i,SEQ_j) > d_X(SEQ_m,SEQ_n) \\
& d_{\text{purpose}}(SQI_i, SQI_j) \neq d_{\text{purpose}}(SQI_m, SQI_n) \\
\text{or} \\
& d_X(SEQ_i,SEQ_j) < d_X(SEQ_m,SEQ_n) \\
& d_{\text{purpose}}(SQI_i, SQI_j) < d_{\text{purpose}}(SQI_m, SQI_n) \\
\}
\]

\[ N = \text{number of simulations} \]
2) Separation value.

\[ d_X \]

\[ H \]

\[ K \]

Simulation couples \((i,j)\)

Acceptance criteria

Fraction of points in the upper left or lower right quadrant

\[
m_{ij} = \begin{cases} 
1 & \iff \left[ d_{\text{purpose}}(SOI_i, SOI_j) - H \right] \cdot \left[ d_{X_{\text{best}}}(SEQ_i, SEQ_j) - K \right] > 0 \\
0 & \text{else} 
\end{cases}
\]

\[
s(K) = \frac{\sum_{ij} m_{ij}}{N(N-1)}. 
\]
Example based on MUST simulations for COST732

Array of obstacles – wind tunnel

- Point release at ground level
- Concentration measurements at H/2 (H=height of the obstacle).
- Flow measurements (velocity components, TKE).

17 simulations (different models, different users, different set-ups)

<table>
<thead>
<tr>
<th>Model</th>
<th>Developer</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>FINFLO</td>
<td>Helsinki University of Technology, Finland</td>
<td>Hellstein (3 sim.)</td>
</tr>
<tr>
<td>FLUENT</td>
<td>ANSYS (commercial code)</td>
<td>Franke, Goricsan (2 sim.), Santiago, Buccolieri.</td>
</tr>
<tr>
<td>M2UE</td>
<td>Tomsk State University, Russia, and Danish Meteorological Institute</td>
<td>Nuterman, Starchenko and Baklanov</td>
</tr>
<tr>
<td>MISKAM</td>
<td>University of Mainz, Germany</td>
<td>Ketzel (2 sim.), Goricsan (3 sim.)</td>
</tr>
<tr>
<td>STAR CD</td>
<td>CD-ADAPCO (commercial code)</td>
<td>Brzozwski</td>
</tr>
<tr>
<td>VADIS</td>
<td>University of Aveiro, Portugal</td>
<td>Costa and Tavares</td>
</tr>
<tr>
<td>ADREA</td>
<td>Environmental Research Laboratory of NCSR “Demokritos”, Greece</td>
<td>Efthimiou and Bartzis</td>
</tr>
</tbody>
</table>
To test the methodology we need a case where both $d_{\text{purpose}}$ and $d_X$ can be computed.

So let assume:

$$d_{\text{purpose}}(SQI_i, SQI_j) = 2 \frac{\max(C_i(x)) - \max(C_j(x))}{\max(C_i(x)) + \max(C_j(x))}$$

Relative difference of maximum of concentration at the measurement points.

**Candidate $d_X$ (not involving concentration measurements)**

$$d_{hrvv}(SEQ_i, SEQ_j) = 1 - \text{HitRate}(\text{vect}_i, \text{vect}_j)$$  
Horizontal velocity

$$d_{hrdd}(SEQ_i, SEQ_j) = 1 - \text{HitRate}(\text{dir}_i, \text{dir}_j)$$  
Wind direction

$$d_{hrvx}(SEQ_i, SEQ_j) = 1 - \text{HitRate}(\text{vx}_i, \text{vx}_j)$$  
X-velocity (from profiles)

$$d_{hrvz}(SEQ_i, SEQ_j) = 1 - \text{HitRate}(\text{vz}_i, \text{vz}_j)$$  
Vertical velocity (from profiles)

$$d_{hrke}(SEQ_i, SEQ_j) = 1 - \text{HitRate}(\text{tke}_i, \text{tke}_j)$$  
TKE

$$d_{hrkez}(SEQ_i, SEQ_j) = 1 - \text{HitRate}(\text{tkez}_i, \text{tkez}_j)$$  
TKE from profiles

$H=0.5$
The highest the value of $\tau$, the most similar are the rankings.
Separation value

Horizontal wind

X-comp. profile

\[ d_{hrvv} \]

\[ d_{hrvxz} \]

TKE

\[ d_{hrtke} \]

\[ d_{purpose} \]

Simulation to simulation

Simulation to observation

\[ \diamond \]

\[ \star \]

\[ H=0.5 \]

\[ K_{best} \]

\[ s(K_{best}) \]

\[ s(K_{best}, Obs) \]

\[ d_{hrvv} \]

\[ 0.34 \]

\[ 0.77 \]

\[ 0.70 \]

\[ d_{hrtke} \]

\[ 0.77 \]

\[ 0.70 \]

\[ 0.65 \]
Conclusions

$\textit{d}_\text{purpose}$ and the acceptance criteria $H$ depend only on the purpose – cannot be computed directly.
Conclusions

- $d_{\text{purpose}}$ and the acceptance criteria $H$ depend only on the purpose – cannot be computed directly.

- $d_X$ and $K_{\text{best}}$ (those actually used) depend, on the purpose, the specific case, the distribution and type of measurements available – can be computed.
Conclusions

\[ d_{\text{purpose}} \] and the acceptance criteria \( H \) depend only on the purpose – cannot be computed directly.

**Methodology**

\[ d_X \] and \( K_{\text{best}} \) (those actually used) depend, on the purpose, the specific case, the distribution and type of measurements available – can be computed.
Conclusions

\( d_{\text{purpose}} \) and the acceptance criteria \( H \) depend only on the purpose – cannot be computed directly.

Methodology

\( d_X \) and \( K_{\text{best}} \) (those actually used) depend, on the purpose, the specific case, the distribution and type of measurements available – can be computed.