Air traffic complexity

A growing need of complexity measures, and a multitude of metrics...
...used for various purposes:

- assessment of ATC/ATFM performances,
- benchmarking new airspace structures,
- design of new tools...

There is no single universal complexity measure.

All we may expect is a set of complexity metrics, useful in a particular context, for a particular purpose.

⇒ we must assess the relevance of the complexity measures to the final goal!
Context and motivation of our study

Context

- Air traffic complexity in relation with the controller’s workload.
- Geographic area: a set of control sectors.
- Time scale: one measure every minute.

What for?

- Purpose: find a criterion to evaluate the controller’s workload.
- Application: optimal sector configurations (pre-tactical level).
  → use relevant complexity metrics instead of incoming flows / sector capacities.
Several studies tried to find a relationship between the complexity metrics and the controller’s workload, using:

- linear or logistic regression,
- cross-sectional time series analysis,
- neural networks...

Existing evaluations of the controller’s workload:

- interactions with the computer,
- physiological activity,
- subjective ratings...

The data:

- may be noisy or biased.
- is difficult to collect (heavy experimental setup),
- and is not available in large quantities.
Our proposal

The sector status (merged, armed, or split) is representative of the controller’s workload.

The aim:
Select a subset of relevant metrics, allowing to predict the sector status.

The data (radar tracks, sector configuration records)
- may still be noisy or biased:
  merging of splitting a sector may have various causes, other than the traffic load,
- is noticeably cheap to collect (ATCC databases),
- and available in large quantities.
- It reflects an objective operational reality.
Evaluation of Complexity Metrics:

**Metrics evaluation method**

**Principal Component Analysis**
- Identify the correlations among the metrics,
- Reduce the dimensionality of the problem.

**Select the most relevant components**
- Apply neural networks:
  - input: main Components (eigenvalue > 1)
  - output: prediction of the sector status
- Select model with best fit: predicted vs. recorded status.

**Evaluate the individual metrics**
- Consider the metrics related to each relevant component
- Select the most relevant subset in each component
Neural networks with one hidden layer

Input vector $x = (x_1, ..., x_i, ..., x_p)^T$
Output vector $y = (y_1, ..., y_k, ..., y_q)^T$

$$y_k = \psi\left(\sum_{j=1}^{q} w_{jk} \Phi\left(\sum_{i=1}^{p} w_{ij} x_i + w_{0j}\right) + w_{0k}\right) \quad (1)$$

$$\Phi(z) = \frac{1}{1 + e^{-z}} \quad \psi(z_k) = \frac{e^{z_k}}{\sum_{m=1}^{C} e^{z_m}}$$
Neural network’s training

Classification problem –> assign each input vector $x$ to a class
merged, armed, or split

Training

- Choose the weights and biases so as to minimize the cross-entropy:

$$E(w) = - \sum_{n=1}^{N} \sum_{k=1}^{C} t_k^{(n)} \ln(y_k^{(n)})$$

- considering $N$ data samples (train set), and $C = 3$ classes,
- where $t$ is the target vector (known class vector)
  - merged class: $t = (1, 0, 0)^T$
  - armed class: $t = (0, 1, 0)^T$
  - split class: $t = (0, 0, 1)^T$
Software and input data

**Software**
- package nnet (written by Pr Ripley) of the R environment,
- training: quasi-Newton minimization method (BFGS).

**Data**
- recorded radar tracks and sector configurations of the 5 French ATCCs. One day of traffic (1st June, 2003),
- 103 different sectors (elementary or group) were armed,
- metrics computed every round minute of the day.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Merged</th>
<th>Armed</th>
<th>Split</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>71270</td>
<td>46.6%</td>
<td>27.0%</td>
<td>26.4%</td>
</tr>
<tr>
<td>Test</td>
<td>47513</td>
<td>46.4%</td>
<td>27.0%</td>
<td>26.6%</td>
</tr>
</tbody>
</table>
Building a set of complexity metrics

27 metrics from various sources
(whenever a clear definition was found):
- Chatterji & Sridhar 2001
- Delahaye & Puechmorel 2000
- Performance Review Unit
- P. Averty (CREED)
- Incoming flows (time horizon 5 to 60 minutes)
  + Sector volume $V$

- Individual indicators typically focus on a single aspect of traffic complexity $\Rightarrow$ need to use complementary indicators.
- Many redundancies $\Rightarrow$ need to reduce dimensionality.
The principal component analysis

The standard deviations are:

\[
\begin{array}{cccccccc}
0 & 0.5 & 1.0 & 1.5 & 2.0 & 2.5 & 3.0 & 3.5 \\
\end{array}
\]

- \(6\) components carry a variance greater than one.
The principal component analysis

The first component is a "size factor" $\sim$ number of aircraft. It explains 46.7% of the variance of the sample.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Other components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speed var. &amp; alt. changes</td>
</tr>
<tr>
<td>avg_vs</td>
<td>0.70</td>
</tr>
<tr>
<td>$\sigma_{gs}/\bar{gs}$</td>
<td>0.69</td>
</tr>
<tr>
<td>$\sigma_{gs}^2$</td>
<td>0.68</td>
</tr>
<tr>
<td>Inc. flows</td>
<td>0.60 to 0.67</td>
</tr>
<tr>
<td>Nb</td>
<td>0.47</td>
</tr>
<tr>
<td>$\text{inter} _\text{hori}$</td>
<td>0.44</td>
</tr>
<tr>
<td>$Nb^2$</td>
<td>0.39</td>
</tr>
<tr>
<td>$\text{incen} _\text{c}$</td>
<td>0.32</td>
</tr>
<tr>
<td>$\text{insen} _\text{d}$</td>
<td>0.48</td>
</tr>
<tr>
<td>$\text{div} _\text{p}$</td>
<td>0.37</td>
</tr>
<tr>
<td>$\text{vprox} _\text{2}$</td>
<td>0.70</td>
</tr>
<tr>
<td>$\text{vprox} _\text{1}$</td>
<td>0.51</td>
</tr>
<tr>
<td>% Variance Explained</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>3.8</td>
</tr>
</tbody>
</table>
Neural networks applied to the components

Iterative approach:
- We first use component $C_1$ as sole input,
- then add the sector volume $V$,
- then component $C_2$ to $C_6$, successively,
- At last, we use the 27 components and the volume as input.

**Schwartz Information Criterion:**

\[ BIC = 2\lambda \cdot \ln(N) - 2\ln(L) \]

- $\lambda$ is the number of unadjusted parameters of the model $\rightarrow$ weights and biases,
- $N$ is the size of the data set,
- $\ln(L)$ is the log-likelihood $\rightarrow$ cross-entropy in our case.
Evolution of the BIC for the components

Input variables:
- C1
- V+C1
- V+C1C2
- V+C1to3
- V+C1to4
- V+C1to5
- V+C1to6
- V+AllC

Train/Test

BIC avg

D. Gianazza (LOG), K. Guittet (LEEA) - DSNA-DTI-SDER
Selection and evaluation of air traffic complexity metrics
The network’s results (classification rates, cross-entropy) evolve with:

- the relevant/non-relevant variables added to the model,
- but also the *number of model’s parameters* (weights, biases).

$\Rightarrow$ The BIC is more reliable. Best model: $\{V; C_1; \ldots; C_4\}$

<table>
<thead>
<tr>
<th>Set</th>
<th>Hidden units</th>
<th>Nb params.</th>
<th>Classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>${V; C_1; \ldots; C_{27}}_{test}$</td>
<td>15</td>
<td>483</td>
<td>82.67%</td>
</tr>
<tr>
<td>${V; C_1; \ldots; C_4}_{test}$</td>
<td>15</td>
<td>138</td>
<td>81.82%</td>
</tr>
<tr>
<td>${V; C_1; \ldots; C_4}_{test}$</td>
<td>53</td>
<td>480</td>
<td>83.84%</td>
</tr>
</tbody>
</table>
## Results on the components

Selection process: 27 metrics $\rightarrow$ 4 relevant components

<table>
<thead>
<tr>
<th>Best components model among the candidate models:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$: size factor, number of aircraft,</td>
</tr>
<tr>
<td>$V$: sector volume,</td>
</tr>
<tr>
<td>$C_2$: speed variance, altitude changes,</td>
</tr>
<tr>
<td>$C_3$: incoming flows,</td>
</tr>
<tr>
<td>$C_4$: flows convergence, anticipation of conflicts.</td>
</tr>
</tbody>
</table>

Functional relationship:

- metrics $\rightarrow$ components
- relevant components $\rightarrow$ sector status.
Results on the indicators

For practical purposes, we would prefer:

- relevant metrics $\rightarrow$ sector status

$\implies$ Evaluate the metrics of each relevant component.

<table>
<thead>
<tr>
<th>Component</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_2$</td>
<td>avg_vs, $\frac{\sigma_{gs}}{gs}$, $\sigma_{gs}^2$, $N_{cl}$, $N_{ds}$</td>
</tr>
<tr>
<td>$C_3$</td>
<td>$F_5$, $F_{15}$, $F_{30}$, $F_{60}$</td>
</tr>
<tr>
<td>$C_4$</td>
<td>inter_hori, insen_c, Conv, creed_ok, creed_pb</td>
</tr>
</tbody>
</table>
Classification rates

Correct classification rates

- with the most relevant indicators
  \{V, Nb, \text{avg}\_\text{vs}, F_{60}, F_{15}, \text{inter}\_\text{hori}\}
- using a neural network with 15 hidden units

<table>
<thead>
<tr>
<th>Set</th>
<th>Global</th>
<th>Merged</th>
<th>Armed</th>
<th>Split</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>83.47%</td>
<td>89.07%</td>
<td>65.58%</td>
<td>91.91%</td>
</tr>
<tr>
<td>Test</td>
<td>83.24%</td>
<td>89.14%</td>
<td>64.56%</td>
<td>91.75%</td>
</tr>
</tbody>
</table>
Discussion $C_2$ indicators

$C_2$: ground speed variance and altitude changes

- the component is relevant,
- this contradicts previous works by [Malasonis, Callaham, Wanke, 2003]
  -> nature of data?
  relationship sector volume / type of traffic?
- the average vertical speed $\text{avg} \_ \text{vs}$ is the most relevant in $C_2$
- but $\frac{\sigma_{gs}}{gs}$ and $\sigma_{gs}^2$ gave fairly close results

The average vertical speed is a dynamic indicator
Could it be replaced by a static sector typology, like: $\text{en route}$, $\text{arrival}$, $\text{departure}$, $\text{mixed}$?
-> classify the sectors according to $\text{avg} \_ \text{vs}$. 

D. Gianazza (LOG), K. Guittet (LEEA) - DSNA-DTI-SDER Selection and evaluation of air traffic complexity metrics
Discussion $C_3$ indicators

$C_3$: incoming flows

- $F_5$ gave the worse results,
- $F_{15}$ is relevant
  $\implies$ consistent with working methods: strips distributed about 15 minutes before sector entry,
- $F_{60}$ came as a surprise
  $\implies$ smoothing effect? compensate the variations of the other variables?
Discussion \( C_4 \) indicators

\( C_4 \): flows convergence and anticipation of conflicts

- **inter\_hori** (number of aircraft pairs with crossing angle greater than 20 degrees) is the most relevant,
- other "convergence" indicators also performed well,
- the indicators related to the anticipation of conflicts
  - between converging aircraft
  - in the horizontal plane

perform well, but do not improve the results

\( \rightarrow \) too much correlated to the airspace structure?

More general conflict detection indicators may be useful
(not only converging aircraft, not only in horizontal plane)
Conclusion

27 metrics + sector volume $\longrightarrow$ 6 relevant indicators

Best model for the indicators:
- $V$: sector volume,
- $Nb$: number of aircraft,
- $avg\_vs$: average vertical speed,
- $F_{15}$, $F_{60}$: incoming flows,
- $inter\_hori$: Number of aircraft pairs crossing with an angle greater than 20 degrees.

Not an absolute truth!
- combined effects of the different indicators
- adding or removing some metrics may change the picture.
Neural networks proved efficient.

Consistent results on \textit{train} and \textit{test} data \implies ability to generalize on fresh inputs

Output: probability to belong to a class \textit{(merged, armed, or split)}.

Simple relationship:
\[
\{ V, Nb, avg\_vs, F_{60}, F_{15}, inter\_hori \} \longrightarrow \text{sector status}
\]
\[
y_k = \Psi \left( \sum_{j=1}^{q} w_{jk} \Phi \left( \sum_{i=1}^{p} w_{ij} x_i + w_0j \right) + w_{0k} \right)
\]
with the weights of the trained network.
Conclusion

About 83% of correct classifications, with the best model and with 15 hidden units.

Classification rates or cross-entropy are not good criteria for model selection.

Selection criterion

- The Schwartz’s *Bayesian Information Criterion* allowed model’s selection,
- not biased by the increasing number of weights in the NN.
Dynamic indicators: high variations in time
Sector status: few changes

Research options

- Could static indicators replace dynamic metrics:
  - average vertical speed $\rightarrow$ sector class (en-route, arrival,...),
  - pairs of crossing aircraft $\rightarrow$ sector class (crossing / parallel)?
- Use "smoothed" metrics?
- Time series processing?

Other methods could be tested
(logistic regression, dynamic discrete choice models).
Application: optimal sector configurations

Use air traffic complexity metrics instead of incoming flows / sector capacities.
End of presentation...

... any questions?