

Selection and evaluation of air traffic complexity metrics

D. Gianazza (LOG), K. Guittet (LEEA) - DSNA-DTI-SDER

25th DASC

October 15-19, 2006

- 1 Introduction
- 2 Neural networks
- 3 Complexity Metrics
- 4 Results
- 5 Conclusion

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.

Other works

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.

Metrics evaluation method

Principal Component Analysis

- Identify the correlations among the metrics,
- Reduce the dimensionality of the problem.

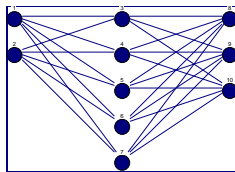
Select the most relevant components

- 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 $x = (x_1, \dots, x_i, \dots, x_p)^T$

Output vector $y = (y_1, \dots, y_k, \dots, y_q)^T$

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

$$\Phi(z) = \frac{1}{1 + e^{-z}} \quad \Psi(z_k) = \frac{e^{z_k}}{\sum_{m=1}^C e^{z_m}}$$

Neural network's training

Classification problem \rightarrow 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.

	Total	Merged	Armed	Split
Train	71270	46.6%	27.0%	26.4%
Test	47513	46.4%	27.0%	26.6%

Building a set of complexity metrics

27 metrics from various sources

(whenever a clear definition was found) :

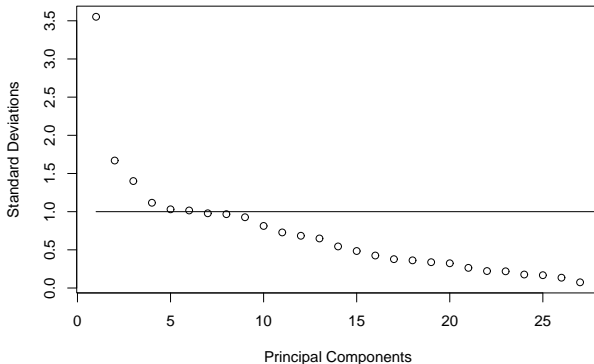
- [Kopardekar & Magyarits (2003)]
- [Chatterji & Sridhar 2001]
- [Delahaye & Puechmorel 2000]
- Performance Review Unit
- P. Averty (CREED)
- Incoming flows (time horizon 5 to 60 minutes)

+ 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 :



⇒ 6 components carry a variance greater than one.

The principal component analysis

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

Indicator	Other components				
	Speed var. & alt. changes	Incoming flows	Structure & Convergence	Divergence	Vertical separation
<i>avg_vs</i>	0.70				
$\sigma_{gs}/\bar{g}s$	0.69				
σ_{gs}^2	0.68				
<i>Inc. flows</i>		0.60 to 0.67			
<i>Nb</i>			0.47		
<i>inter_hori</i>			0.44		
Nb^2			0.39		
<i>incen_c</i>			0.32		
<i>insen_d</i>				0.48	
<i>div_p</i>				0.37	
<i>vprox_2</i>					0.70
<i>vprox_1</i>					0.51
% Variance Explained	10.3	7.3	4.6	3.9	3.8

Neural networks applied to the components

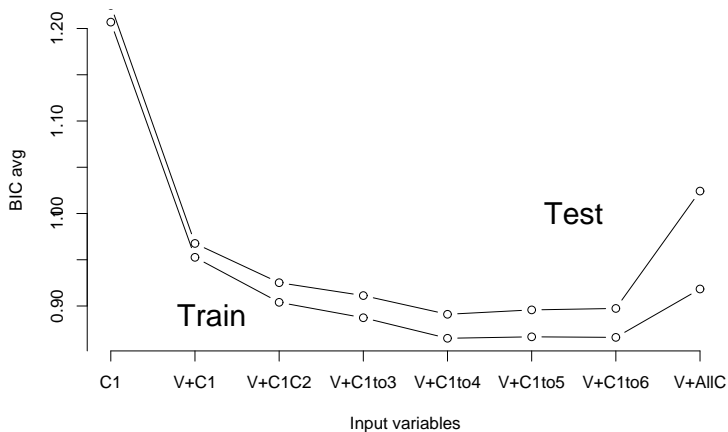
Iterative approach :

- 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 : ($BIC = 2\lambda \cdot \ln(N) - 2\ln(L)$)

- λ is the number of unadjusted parameters of the model
→ weights and biases,
- N is the size of the data set,
- $\ln(L)$ is the log-likelihood → cross-entropy in our case.

Evolution of the BIC for the components



Why use the BIC ?

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).

==> The BIC is more reliable. Best model : $\{V; C_1; \dots; C_4\}$

Set	Hidden units	Nb params.	Classification rate
$\{V; C_1; \dots; C_{27}\}_{test}$	15	483	82.67%
$\{V; C_1; \dots; C_4\}_{test}$	15	138	81.82%%
$\{V; C_1; \dots; C_4\}_{test}$	53	480	83.84%

Results on the components

Selection process : 27 metrics \rightarrow 4 relevant components

Best components model among the candidate models :

- C_1 : size factor, number of aircraft,
- V : sector volume,
- C_2 : speed variance, altitude changes,
- C_3 : incoming flows,
- C_4 : flows convergence, anticipation of conflicts.

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.

Component	Metrics
C_2	avg_vs , $\frac{\sigma_{gs}}{gs}$, σ_{gs}^2 , N_{cl} , N_{ds}
C_3	F_5 , F₁₅ , F_{30} , F₆₀
C_4	inter_hori , <i>insen_c</i> , <i>Conv</i> , <i>creed_ok</i> , <i>creed_pb</i>

Classification rates

Correct classification rates

- with the most relevant indicators
 $\{V, Nb, avg_vs, F_{60}, F_{15}, inter_hori\}$
- using a neural network with 15 hidden units

Set	Global	Merged	Armed	Split
<i>Train</i>	83.47%	89.07%	65.58%	91.91%
<i>Test</i>	83.24%	89.14%	64.56%	91.75%

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 **avg_vs** is the most relevant in C_2
- but $\frac{\sigma_{gs}}{gs}$ and σ_{gs}^2 gave fairly close results

The average vertical speed is a dynamic indicator

Could it be replaced by a static sector typology, like : *en route*, *arrival*, *departure*, *mixed*?

→ classify the sectors according to **avg_vs**.

Discussion C_3 indicators

C_3 : incoming flows

- F_5 gave the worse results,
- **F_{15}** is relevant
→ consistent with working methods : strips distributed about 15 minutes before sector entry,
- **F_{60}** came as a surprise
→ 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

→ 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.

Conclusion

Method

- Neural networks proved efficient.
- Consistent results on *train* and *test* data
⇒ ability to generalize on fresh inputs
- Output : probability to belong to a class
(*merged*, *armed*, or *split*).

Simple relationship :

$\{V, Nb, avg_vs, F_{60}, F_{15}, inter_hori\} \longrightarrow$ 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.

Further work

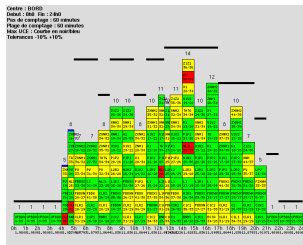
Dynamic indicators : high variations in time
Sector status : few changes

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).

Further work



Application : optimal sector configurations

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

End of presentation...

... any questions ?

