

Analysis of a Workload Model Learned from Past Sector Operations

David Gianazza



Outline

Introduction

Model and method

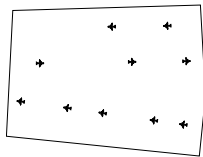
Experimental setup

Results

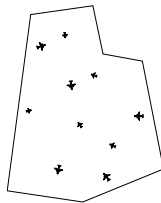
Conclusion

Introduction

ATC complexity



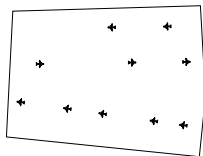
Simple



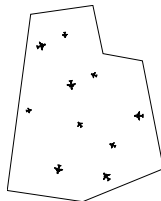
Complex

Introduction

ATC complexity



Simple



Complex

Workload

&

Airspace configuration



Questions related to ATC complexity and workload

Can we predict the workload of ATC controllers, given a traffic situation as input ?

- How to measure the ATC complexity of a traffic situation ?
- How do we quantify workload ?
- Can we build a predictive model of the workload, taking ATC complexity into account ?
- What model should we choose ?

Many answers in the literature

[Mogford et al., 1995, Hilburn, 2004, Delahaye and Puechmorel, 2000, Delahaye and Puechmorel, 2010, Flynn et al., 2005, Welch et al., 2007, Kopardekar and Magyarits, 2003, Chatterji and Sridhar, 1999], etc.

Previous publication at ATM R&D seminar



D. Gianazza.

Learning air traffic controller workload from past sector operations.

In *Proceedings of the 12th USA/Europe Air Traffic Management R & D Seminar*, 2017.

Compared several machine learning methods

- Linear discriminant analysis (LDA),
- Quadratic discriminant analysis (QDA),
- Naive Bayes classifier (NBayes),
- Neural networks (NNet),
- Gradient boosted trees (GBM)

On the workload prediction problem

$$\begin{aligned} \text{workload} &= f(\text{complexity}) \\ \text{complexity} &= g(\text{traffic}, \text{sector}, \text{procedures}) \end{aligned}$$

Previous publication at ATM R&D seminar

ATC complexity metrics taken from the literature

Workload model is learned from past sector operations
considering sector status (collapsed, open, split)

Previous publication at ATM R&D seminar

Experimental setup

- Same ATC sectors for learning and prediction
- Model performance evaluated on different time intervals

Results

- Poor performance of simple methods (linear, quadratic, or naive Bayes)
- Most performant models : NNet,GBM
- About 82% of correct classifications

Subject of this publication

Can we generalize to new sectors ?

New experimental setup :

- Different ATC sectors for learning and prediction
- Model performance evaluated on a same time period but on different sectors

Variations of workload with input ATC complexity values ?

Bagplots of low, normal, high workload data points for all pairs of complexity metrics

Outline

Introduction

Model and method

Experimental setup

Results

Conclusion

Supervised machine learning

Build a statistical model using collected data :

- Measures of complexity metrics extracted from radar data and sector geometry
E.g. : density, incoming flows, climbing/descending a/c, dispersion of speed direction/intensity, vertical velocity variance, number of potential conflicts, sector volume, number of crossing points, etc.
- Workload measures : EEG, physiological parameters, pupil diameter, eye fixation points, duration and content of radio exchanges, subjectives ratings, **past sector configurations**.

Assumption : $workload = f(complexity)$

- True relationship $y = f(x)$ is unknown
- Approximation $y = h(x) + \varepsilon$
- h is tuned on a set of examples $(x_1, y_1), \dots, (x_N, y_N)$

Our approach

Basic assumption :

Decisions to split or merge ATC sectors are caused by workload variations.

Workload measure $\approx$ ATC sector status

- *Low workload* when sector s is collapsed with other sectors
- *Normal workload* when the sector s is operated as is
- *High workload* when s is split into several smaller sectors operated on different working positions

Our approach

Supervised learning

- Unknown relationship $y = f(x)$
between workload y and ATC complexity metrics x
- approximated by $\hat{y} = h(x)$ using examples
 $(x_1, y_1), \dots, (x_N, y_N)$

Regression or classification problems

- Regression problem : $y \in \mathbb{R}$ or $y \in \mathbb{R}^p$
as in [Kopardekar and Magyarits, 2003,
Chatterji and Sridhar, 1999]
- Classification problem : $y \in \{\mathcal{C}_1, \dots, \mathcal{C}_P\}$ is a category

Our approach

Supervised learning

- Unknown relationship $y = f(x)$
between workload y and ATC complexity metrics x
- approximated by $\hat{y} = h(x)$ using examples
 $(x_1, y_1), \dots, (x_N, y_N)$

Regression or classification problems

- ~~Regression problem~~ : ~~$y \in \mathbb{R}$ or $y \in \mathbb{R}^p$~~
as in [Kopardekar and Magyarits, 2003,
Chatterji and Sridhar, 1999]
- **Classification problem** : $y \in \{\mathcal{C}_1, \dots, \mathcal{C}_P\}$ is a category

Here, the output is a category : *low*, *normal* or *high* workload.

Output variable y

Examples extracted from past sector operations :

<i>Collapsed sector</i>	$y = (1, 0, 0)^T$	Underload
<i>Opened sector</i>	$y = (0, 1, 0)^T$	Normal workload
<i>Split sector</i>	$y = (0, 0, 1)^T$	Overload
Other	Not used	unknown workload

Advantages, compared to other workload measures

- Historical data is easy to collect
- More than 510 ATC sectors for the 5 french ATCCs
- Not subjective

Drawback

- Only three levels of workload

Output variable y

Examples extracted from past sector operations :

<i>Collapsed sector</i>	$y = (1, 0, 0)^T$	Underload
<i>Opened sector</i>	$y = (0, 1, 0)^T$	Normal workload
<i>Split sector</i>	$y = (0, 0, 1)^T$	Overload
Other	Not used	unknown workload

Advantages, compared to other workload measures

- Historical data is easy to collect
- More than 510 ATC sectors for the 5 french ATCCs
- Not subjective

Drawback

- Only three levels of workload

Input variables ?

What are the most relevant ATC complexity metrics for our problem ?

Previous research :



D. Gianazza and K. Guittet.

Evaluation of air traffic complexity metrics using neural networks and sector status.

In Proceedings of the 2nd International Conference on Research in Air Transportation. ICRAT, 2006.



D. Gianazza and K. Guittet.

Selection and evaluation of air traffic complexity metrics.

In Proceedings of the 25th Digital Avionics Systems Conference. DASC, 2006.



D. Gianazza.

Smoothed traffic complexity metrics for airspace configuration schedules.

In Proceedings of the 3rd International Conference on Research in Air Transportation. ICRAT, 2008.

Input variables

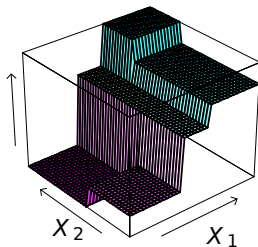
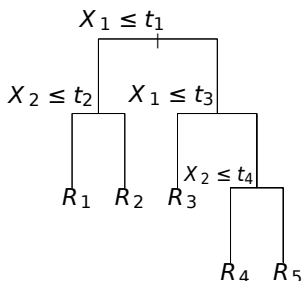
Among 27 ATC complexity metrics from the literature, we found 6 most relevant variables for our problem :

- *vol*, the airspace volume
- *nb*, the number of aircraft
- *flow15*, the incoming traffic flow within the next 15 minutes
- *flow60*, the incoming traffic flow within a 1 hour time horizon
- *avg_vs*, the average absolute vertical speed of the aircraft within the sector,
- *inter_hori*, the number of speed vector intersections with an angle greater than 20 degrees.

Computed from radar tracks and sector geometries

Classification and regression trees

Elements of Statistical Learning (2nd Ed.) c Hastie, Tibshirani & Friedman 2009 Chap 9



$$h_T(\mathbf{x}) = \sum_{j=1}^{|T|} c_j(\mathbf{x}) \mathbb{1}_{R_j}(\mathbf{x}) \quad (1)$$

Regression with quadratic loss : $c_j(\mathbf{x}) = \frac{1}{N_j} \sum_{\mathbf{x}_n \in R_j} y_n$

Classification : $c_j(\mathbf{x}) = \mathcal{C}_p$ where $\mathcal{C}_p$ is the most occurring class in R_j

Gradient-boosted trees

Sum of tree models, iteratively adjusted.

Model update :

$$h_m : \mathbf{x} \rightarrow h_{m-1}(\mathbf{x}) + \nu \sum_{R_j \in T_m} \gamma_{mj} \mathbb{1}_{R_j}(\mathbf{x})$$

Parameters :

- J : tree size
Usually, $4 \leq J \leq 8$
- M : number of boosting iterations
Can be selected by an *early stopping* procedure
- ν : shrinkage parameter
Usually, $\nu \leq 0.1$

Outline

Introduction

Model and method

Experimental setup

Results

Conclusion

Dataset

Historical data from 5 french ATC centers (2 weeks, Oct. 2016) :

- Radar tracks
- Past sector operations
- Number of samples balanced as well as possible, **per sector**, among workload categories having non-zero occurrences
- Completion using sectors that were not opened
- 147878 complexity and workload samples
- 163 elementary sectors, from which only 16 were actually opened
- 369 non-elementary sectors

Experimental setup

Objective : Assess model on a different set of sectors than the ones used to train it

The GBM method requires hyperparameter tuning (J , M , ν)

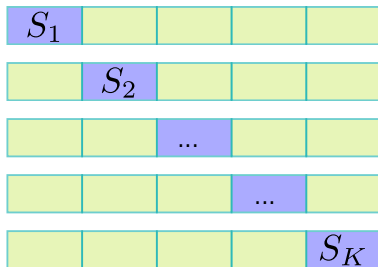
Nested cross-validations :

- 7-fold outer cross-validation to assess model performance
- 5-fold inner cross-validation to select hyperparameter values

K -fold cross-validation

S is split into K subsets

Model selection:



For each candidate model:

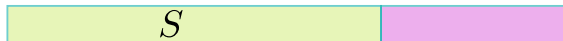
For each fold k :

- Adjust model using $S_{-k} = \bigcup_{i \neq k} S_i$
- Assess error using S_k

Compute the cumulated error for the K folds

Select model minimizing the error on the K validation subsets S_k

Prediction:



Adjust the best model found

Apply to fresh entries

Experimental setup (continued)

Samples concerning a given sector must be in only one subset S_k

Hyperparameter grid for inner cross-validation :

Method	Hyperparameter grid
GBM _(m,J,ν)	$m = \{5000, 6000, 7000\}$ $J = \{2, 3, 4\}$ $\nu = \{1e-4, 5e-4, 1e-3, 1e-2, 1e-1\}$

Model performance evaluation

		Reference		
		C1	C2	C3
Predicted	C1	a	b	c
	C2	d	e	f
	C3	g	h	i

$$Accuracy = \frac{a + e + i}{a + b + c + d + e + f + g + h + i}$$

$$Recall(C1) = \frac{a}{a + d + g}$$

$$Precision(C1) = \frac{a}{a + b + c}$$

Outline

Introduction

Model and method

Experimental setup

Results

Conclusion

Results

All sectors

	Overall	Low	Normal	High
Correct classif.	0.759 (0.03)	0.757 (0.046)	0.736 (0.074)	0.808 (0.100)
Precision	0.771 (0.034)	0.817 (0.065)	0.680 (0.042)	0.815 (0.112)

Results

Elementary sectors

	Overall	Low	Normal	High
Correct classif.	0.82 (0.092)	0.904 (0.062)	0.507 (0.197)	NaN (NA)
Precision	0.519 (NA)	0.856 (0.112)	0.539 (0.271)	0 (NA)

Results

Non-elementary sectors

	Overall	Low	Normal	High
Correct classif.	0.757 (0.029)	0.739 (0.041)	0.748 (0.060)	0.808 (0.100)
Precision	0.769 (0.031)	0.814 (0.068)	0.679 (0.047)	0.815 (0.112)

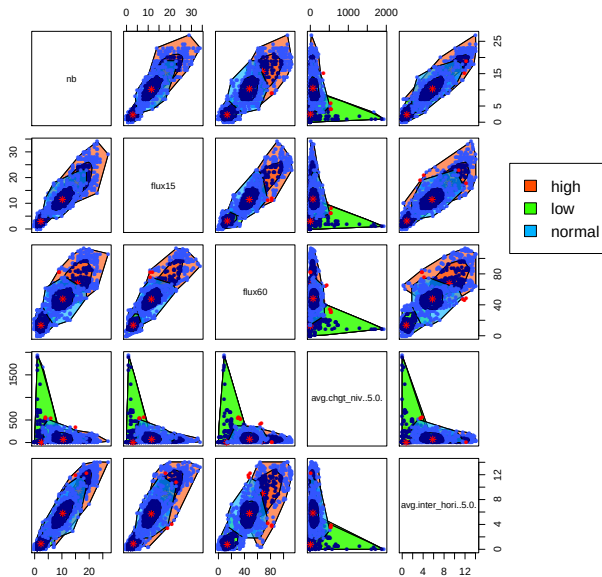
Results

Non-elementary sectors having no instance of the “high workload” class

	Overall	Low	Normal	High
Correct classif.	0.738 (0.033)	0.713 (0.058)	0.760 (0.089)	NaN (NA)
Precision	0.504 (0.020)	0.784 (0.087)	0.729 (0.056)	0.000 (0.000)

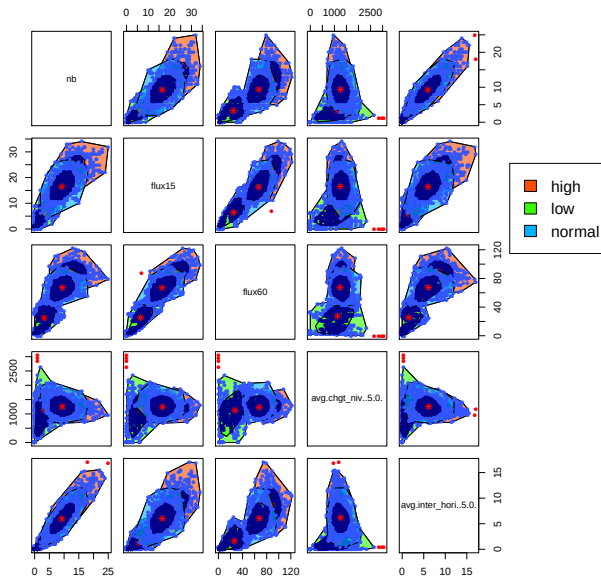
En-route sector N

Sector N



Pre-approach sector AENB

Sector AENB



Outline

Introduction

Model and method

Experimental setup

Results

Conclusion

Conclusion

Model performance with geographic segregation of the training and test sets

- Around 76 % of overall accuracy
- (vs. 82% with temporal segregation)

Model performance on (non-splittable) elementary sectors

- Poor performance for the “normal” class
 - No data for the “high” workload class
 - Few data for the “normal” class
 - No other splittable sectors with similar characteristics

Good performance on non-elementary sectors, including those with no data in the “high” workload class

Conclusion

Model performance with geographic segregation of the training and test sets

- Around 76 % of overall accuracy
- (vs. 82% with temporal segregation)

Model performance on (non-splittable) elementary sectors

- Poor performance for the “normal” class
 - No data for the “high” workload class
 - Few data for the “normal” class
 - No other splittable sectors with similar characteristics

Good performance on non-elementary sectors, including those with no data in the “high” workload class

Conclusion

Model performance with geographic segregation of the training and test sets

- Around 76 % of overall accuracy
- (vs. 82% with temporal segregation)

Model performance on (non-splittable) elementary sectors

- Poor performance for the “normal” class
 - No data for the “high” workload class
 - Few data for the “normal” class
 - No other splittable sectors with similar characteristics

Good performance on non-elementary sectors, including those with no data in the “high” workload class

Conclusion

Can we predict the workload of ATC controllers, given a traffic situation as input ?

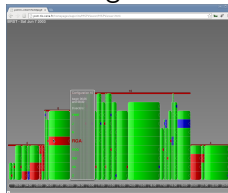
Yes, but...

- Gradient-boosted tree model not usable for elementary sectors
Workaround : Use a simpler model for elementary sectors
- Model captures relationship between workload and complexity + some characteristics of the data
- Only 3 workload levels (low, normal, high) $\Rightarrow$ does not capture fine-grain variations of workload
- In particular, coarse relation between workload and climbs/descents
avg_vs could probably be replaced with a categorical variable with one value per sector

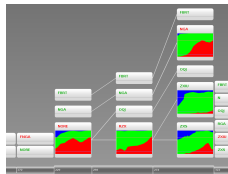
Potential benefits

Improved workload prediction

→ Predict optimal sector configurations



→ Anticipate overloads



→ “What-if” evaluation of envisioned ATFCM measures

Further works

- Effects of seasonality ?
- Use larger datasets (1 year of traffic and sector operations)
- Find optimal sequences of sector configurations, with respect to workload





Chatterji, G. B. and Sridhar, B. (1999).

Neural network based air traffic controller workload prediction.

In *American Control Conference, 1999. Proceedings of the 1999*, volume 4, pages 2620–2624. IEEE.



Delahaye, D. and Puechmorel, S. (2000).

Air traffic complexity : towards intrinsic metrics.

In *Proceedings of the third USA/Europe Air Traffic Management R & D Seminar*.



Delahaye, D. and Puechmorel, S. (2010).

Air traffic complexity based on dynamical systems.

In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 2069–2074. IEEE.



Flynn, G., Benkouar, A., and Christien, R. (2005).

Adaptation of workload model by optimisation algorithms.

Technical report, Eurocontrol.



Hilburn, B. (2004).

Cognitive complexity in air traffic control, a litterature review.

Technical report, Eurocontrol experimental centre.



Kopardekar, P. and Magyarits, S. (2003).

Measurement and prediction of dynamic density.

In *Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar*.



Mogford, R., Guttman, J. A., Morrow, S. L., and Kopardekar, P. (1995).

The complexity construct in air traffic control : A review and synthesis of the literature.

Technical report, FAA Technical Center : Atlantic City.



Welch, J. D., Andrews, J. W., Martin, B. D., and Sridhar, B. (2007).

Macroscopic workload model for estimating en route sector capacity.

In *Proc. of 7th USA/Europe ATM Research and Development Seminar, Barcelona, Spain*.