

# Data-driven modelling and validation of the inbound flow at some major European airports

Carlo Lancia <sup>1</sup> Guglielmo Lulli <sup>2</sup>

<sup>1</sup>Mathematical Institute Leiden University,

<sup>2</sup>Lancaster University Management School

ECSO 2017, August 22, 2017

# Summary

## Problem

Estimating the daily demand at major European airports

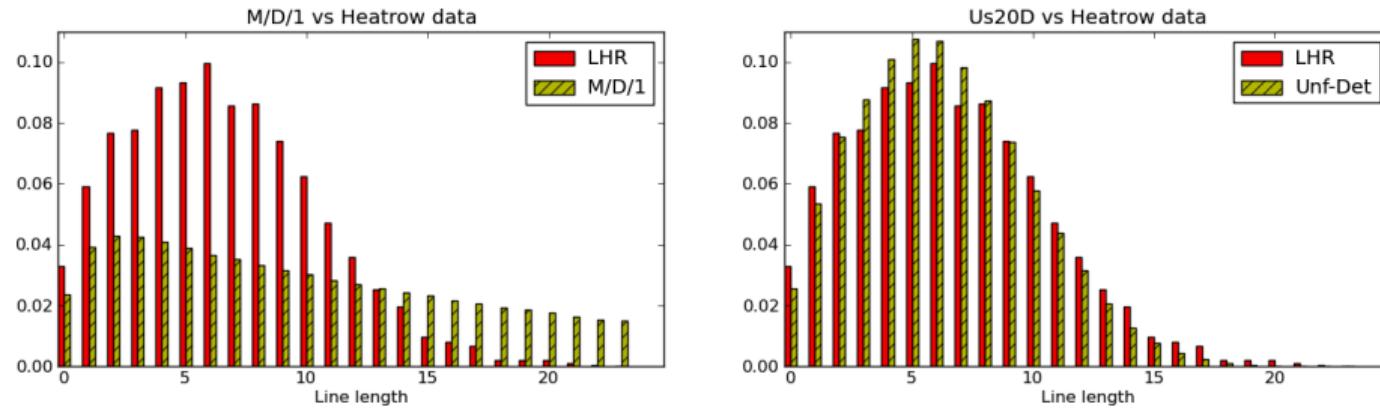
## Data and Methods

- Entrance time in a cylinder of radius 40 NM around airport
- Data-driven definition of Poisson process and Pre-scheduled Random Arrivals
- Comparison of the two processes

## Conclusions

PSRA should be preferred whenever mathematical tractability is not an issue

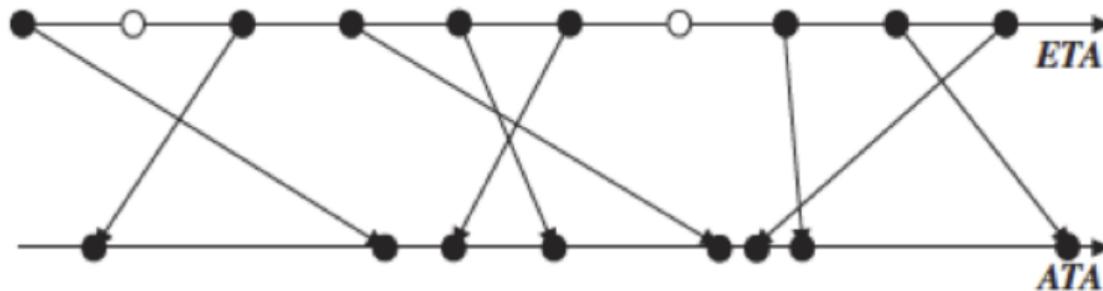
# Modelling the length of inbound queue at Heathrow (EGLL)



From Caccavale, Iovanella, Lancia, Lulli, Scoppola (2014) A model of inbound air traffic: The application to Heathrow airport.

*Journal of Air Transport Management*, 34, 116–122

## Pre-scheduled random arrivals



From Caccavale et al. (2014) *J Air Transp Manag*, 34, 116–122

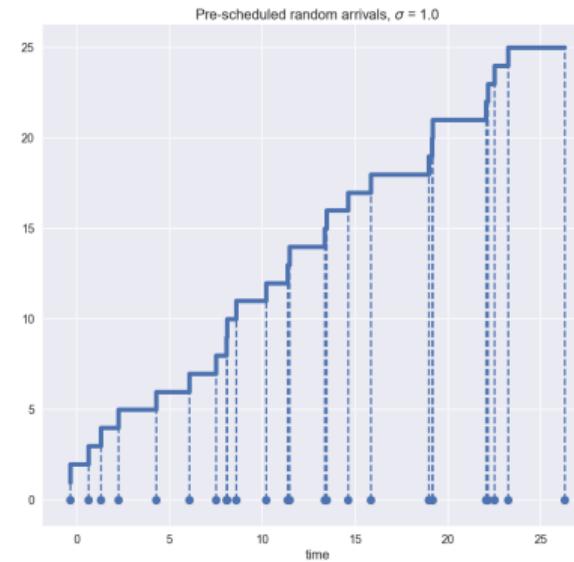
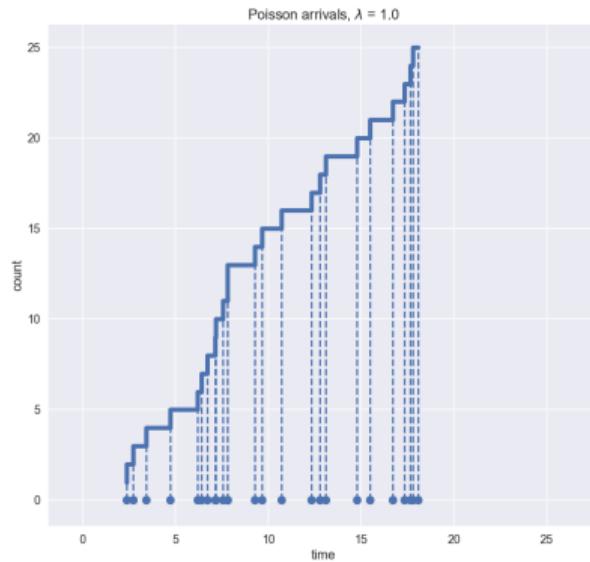
### Idea

Fixed, deterministic schedule perturbed by IID random variables

$$t_i = \frac{i}{\mu} + \xi_i$$

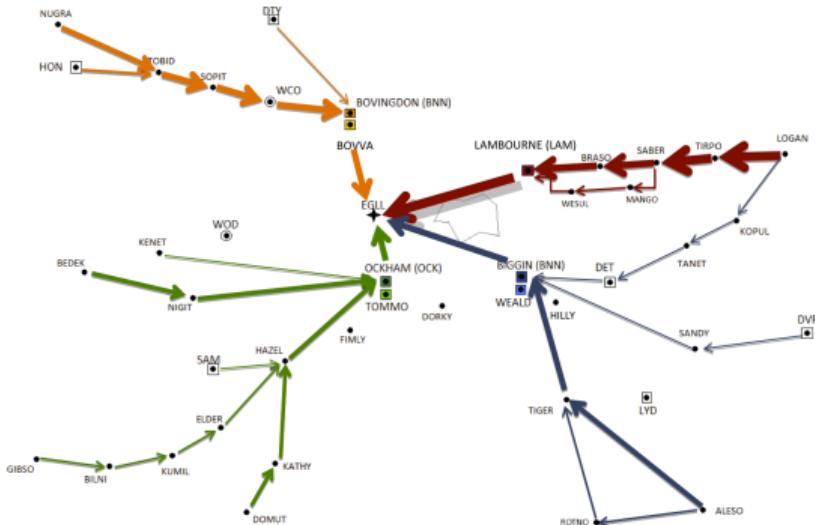
Optional thinning process (hollow dots)

# Poisson and PSRA model



$\cdot$ /D/1 queue model: single server, deterministic service time

# Modelling the length of inbound queue at Heathrow (EGLL)



From Caccavale et al. (2014) *J Air Transp Manag*, 34, 116–122

- Indirect comparison with Poisson, i.e. through a queueing model
- Analysis overlooks action of TC on inbound stream
- *This leads to overestimating PSRA parameter  $\sigma$*
- Deterministic schedule of PSRA is equally spaced in time (this is ok for EGLL, though)

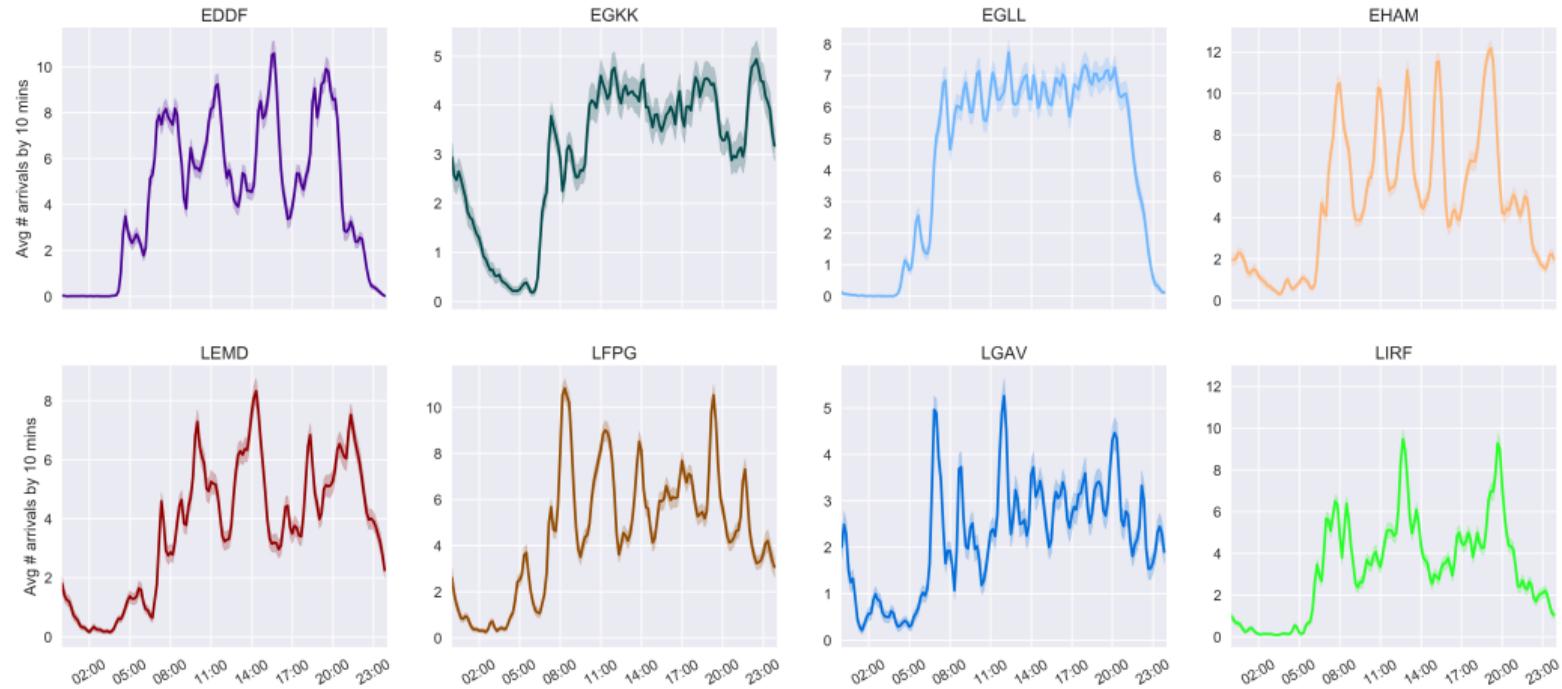
## Dataset overview

airport name	ICAO code	sample size
Frankfurt am Main International Airport	EDDF	58167
London Gatwick Airport	EGKK	39746
London Heathrow Airport	EGLL	56716
Amsterdam Airport Schiphol	EHAM	63279
Madrid Barajas International Airport	LEMD	48162
Charles de Gaulle International Airport	LFPG	60122
Athens International Airport	LGAV	29503
Rome Fiumicino International Airport	LIRF	43333

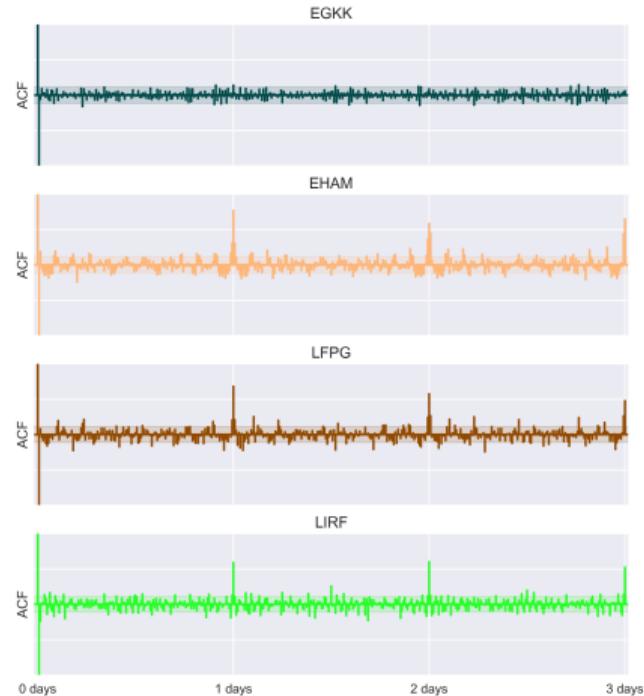
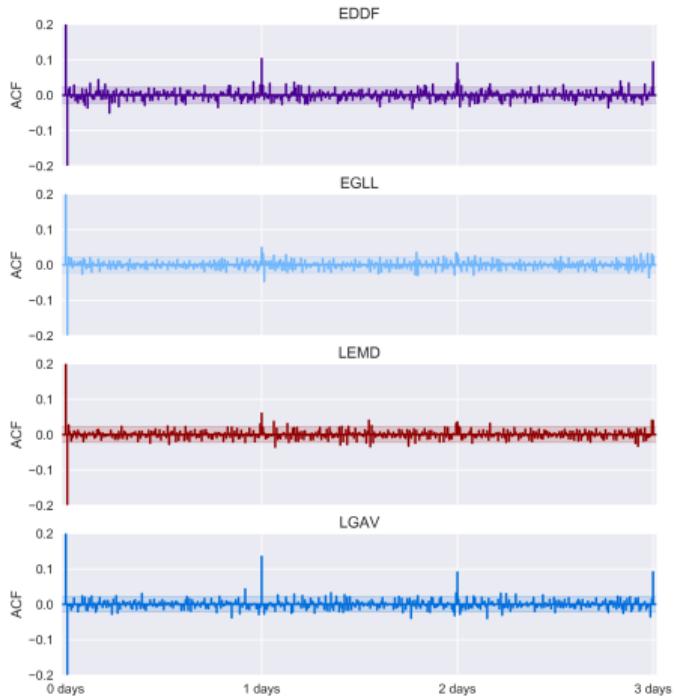
Study period goes from June 15 to September 15, 2016

We model the demand using first time within 40 NM from airport

# Average daily demand



# Daily periodicity

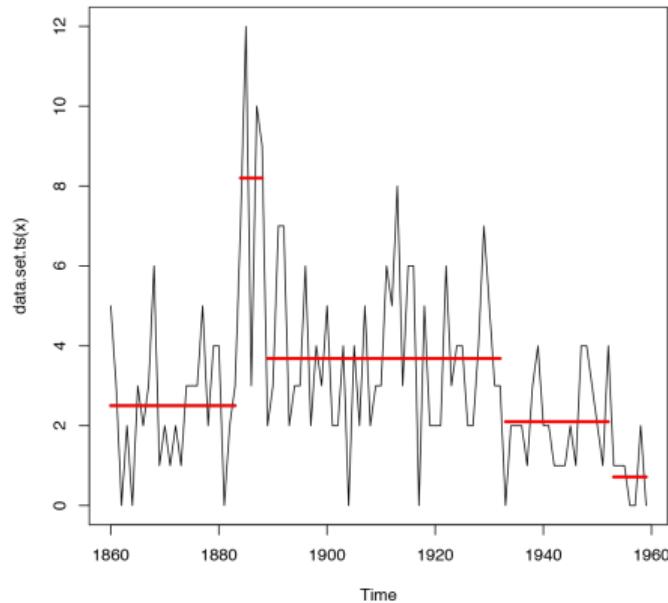


# Data-driven Poisson process

1. Aggregate arrivals TS, e.g. by intervals of 10 minutes
2. Run a **changepoint-detection** algorithm, e.g. PELT, under the null hypothesis of Poissonian arrivals to obtain
  - ▶ changepoint time  $\hat{t}_k$
  - ▶  $\hat{\lambda}_k$ , estimated intensity in  $[\hat{t}_k, \hat{t}_{k+1})$
3. **Cluster** couples  $\{(\hat{t}_k, \hat{\lambda}_k)\}_k$ , e.g. via DBSCAN
4. Compute the **centroid**  $(\bar{t}_i, \bar{\lambda}_i)$  of each cluster
5. Define a **step-wise, periodic intensity function** that takes on  $\bar{\lambda}_i$  for  $t \in [\bar{t}_i, \bar{t}_{i+1})$

# Changepoint detection via PELT

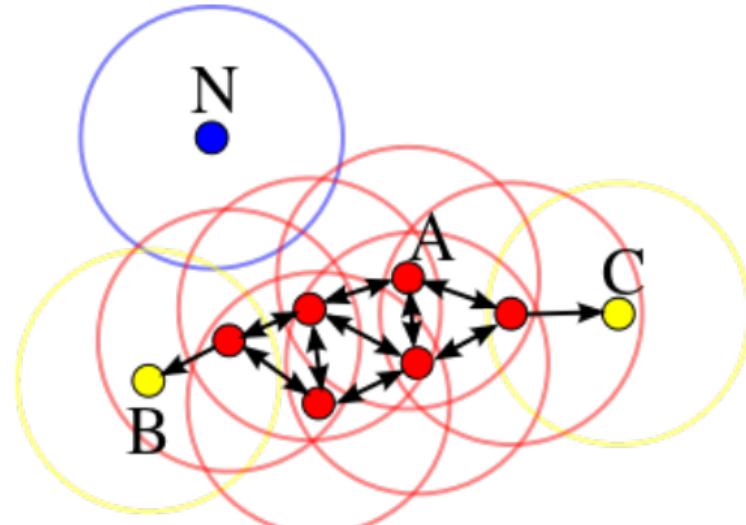
- PELT is an algorithm for multiple changepoint detection in a time series
- It can detect changes in mean, variance, or both
- Software available via R package `changepoint`
- We work under the null of Poisson arrivals



Killick, Eckley (2014) *J Stat Soft* 58(3)

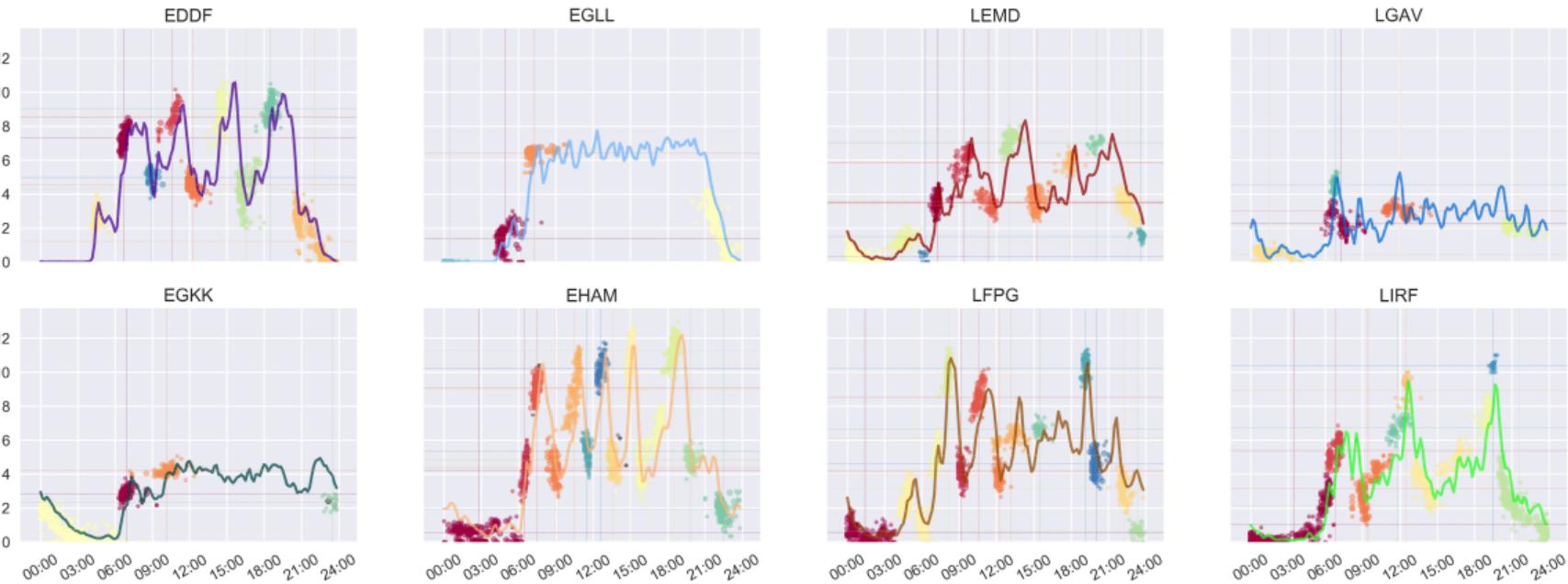
# Clustering via DBSCAN

- Density-based spatial clustering of applications with noise (DBSCAN)
- Two hyperparameters,  $\varepsilon$  and  $n_{pts}$
- Core points (red) have at least  $n_{pts}$  other points within distance  $\varepsilon$
- Non-core (yellow) lie within  $\varepsilon$  from a core point
- The remainders are outliers (blue)



Source: Wikimedia Commons

# Implementation of data-driven Poisson process



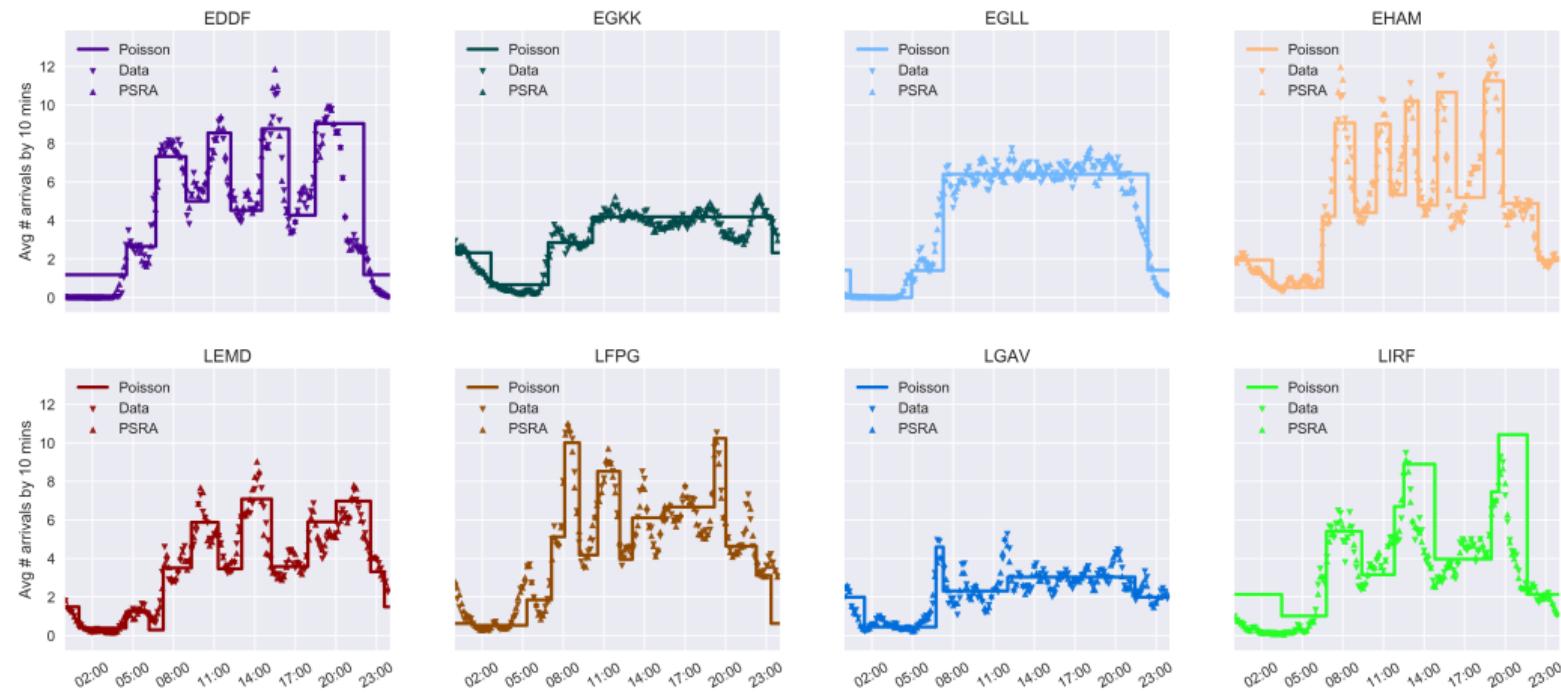
## Data-driven PSRA

1. Let  $t_i^{M3}$  the *actual arrival* time at 40 NM
2. Let  $t_i^{M1}$  the *anticipated arrival* time at 40 NM (from the last flight plan agreed with Eurocontrol)
3. Compute the *delays*  $\delta_i = t_i^{M3} - t_i^{M1}$
4. Define

$$t_i = t_i^{M1} + \xi_i$$

where  $\{\xi_i\}_i$  are IID rvs drawn from the empirical distribution of the delays  $\{\delta_i\}_i$

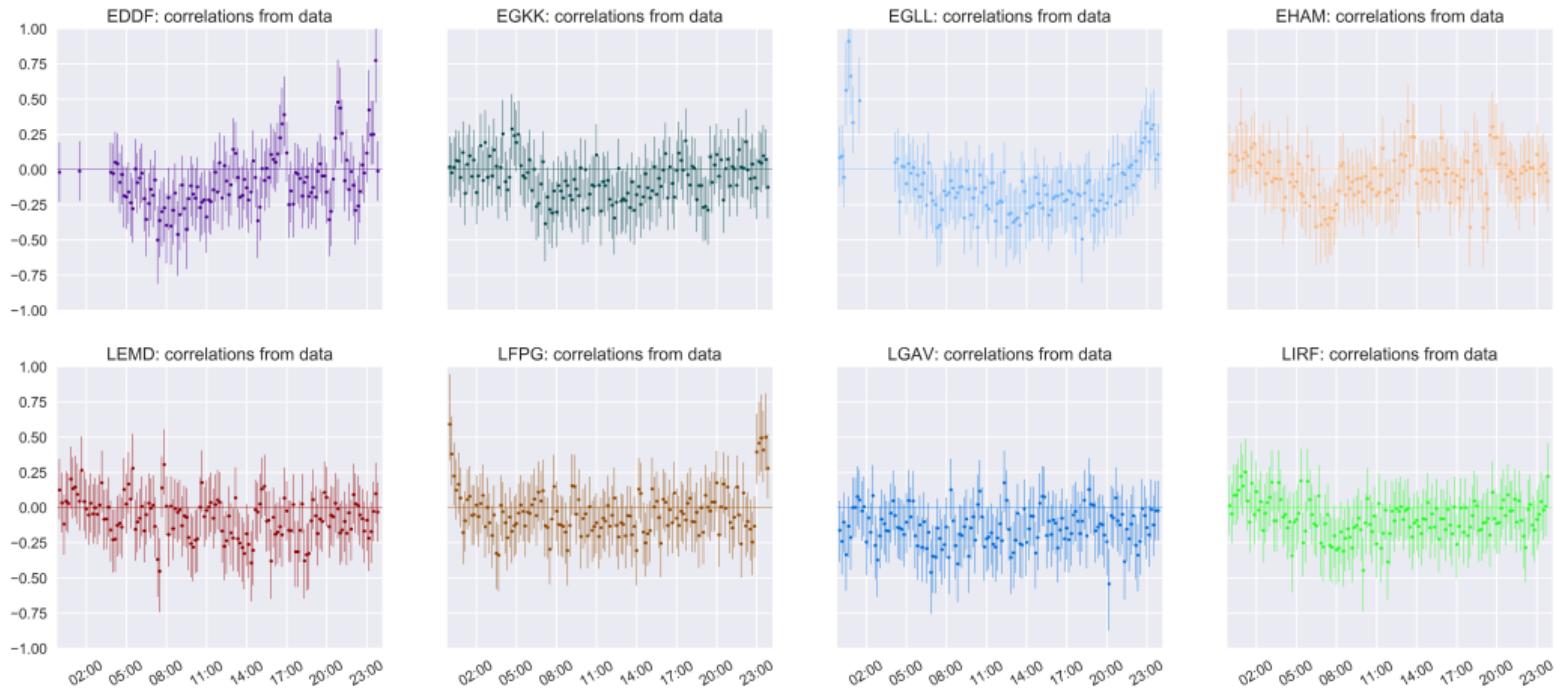
# Data-driven Poisson vs PSRA



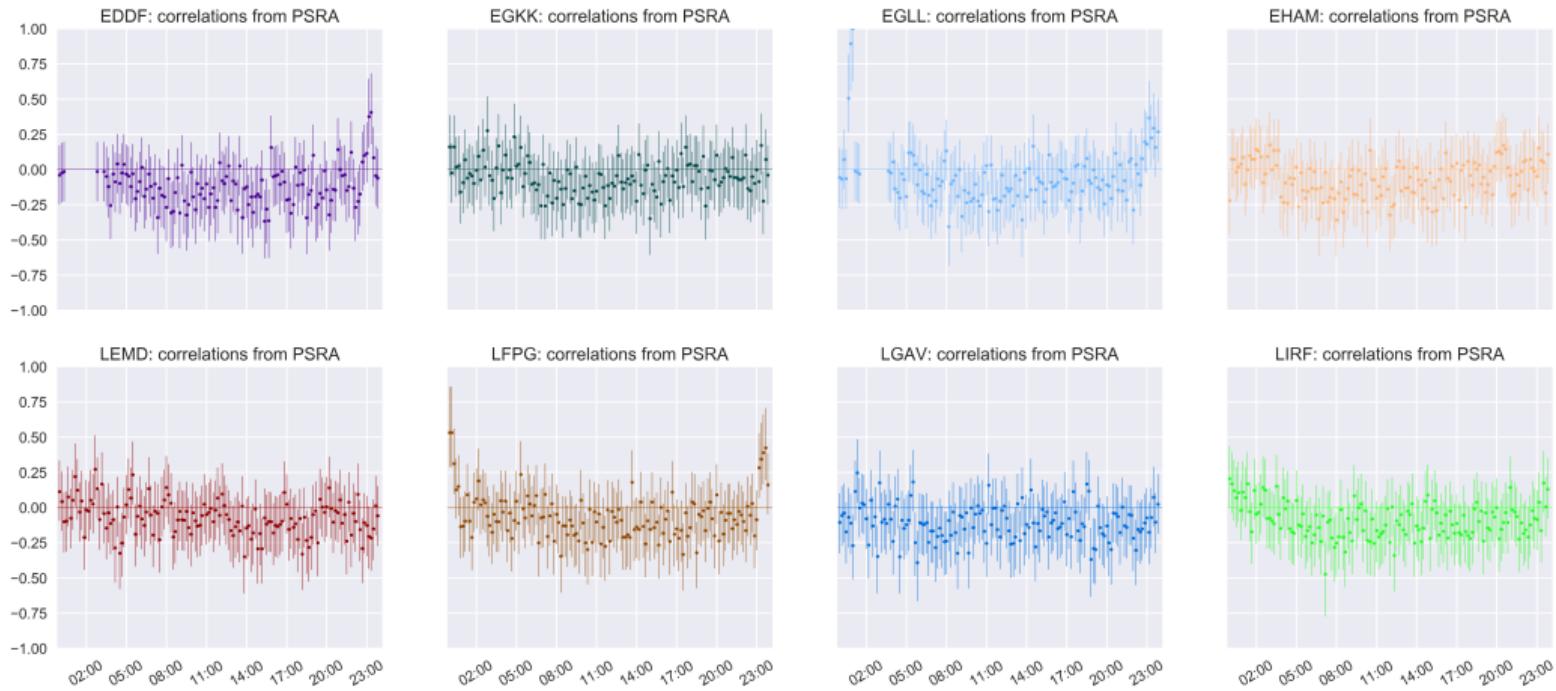
## Data-driven Poisson vs PSRA

- PSRA outperform Poisson arrivals in reproducing the average daily demand
- Average demand reproduced with Poisson model by **forcing** intensity to vary over a finer time scale
  - ▶ **CAVEAT** number of parameters requires increases in a sensible manner
- PSRA is non-parametric
- A semi-parametric model is possible by defining  $\xi_i = \xi_i(Z)$
- PSRA inherits **arrivals correlation-structure** from the M1 schedule
- In contrast, Poissonian arrivals have **independent increments** by definition

# Arrivals correlation



# Correlations from simulation of PSRA



# Conclusions

- PSRA describes better the inbound flow than Poisson arrivals
- Poisson arrivals lead to overestimation of the queue
- Only advantage of Poisson is mathematical tractability
- If time scale is small, only the transient matters, i.e. no benefit
- PSRA are difficult to tackle, but efficient approximation schemes are possible, see Lancia et al (2017) <https://arxiv.org/abs/1302.1999>

## Acknowledgments

Lorenzo Capanna and Luigi De Giovanni for their help with querying the DDR database

## Preprint

Lancia, Lulli (2017) Data-driven modelling and validation of aircraft inbound-stream at some major European airports <https://arxiv.org/abs/1708.02486>