PRC Data Challenge

Method and Results of Team “Likable Jelly”

Richard Alligier, David Gianazza

Ecole Nationale de l’Aviation Civile

PRC Data Challenge

Team "Likable Jelly":

- Richard Alligier, assistant professor at ENAC
- David Gianazza, associate professor at ENAC

PRC data challenge

Develop an open Machine Learning model to predict Aircraft Take-Off Weight (TOW) based on flight and trajectory data.
Provided Files:

- `challenge_set.csv` and `final_submission_set.csv`
  - flight identification:callsign
  - origin/destination: DEParture (ADEP), and DEStination (ADES)
  - timing: date of flight, actual off-block time, arrival time
  - aircraft: aircraft type code
  - airline: (obfuscated) Aircraft Operator code (airline)
  - operational values: flight duration, taxi-out time, flown distance
- OpenSky Network's ADS-B `2022-XX-XX.parquet` files
  - timestamp, latitude, longitude, altitude, groundspeed, ROCD, T and wind

Machine Learning Model: Our Solution

$$\mathrm{TOW} = \mathrm{model}\left(\mathrm{FlightInfo}, \mathrm{WeatherAtAirports}, \mathrm{Trajectory}\right)$$

model

- LightGBM library; an efficient gradient boosted trees library
- Hyper-parameters:
  - Number of boosting iterations (it i.e. number of trees): $50,000$
  - Random search to select size of the trees and regularization parameters

Input Variables

611 input variables extracted from different sources

Input Variables

$$\text{TOW} = \text{model}\,(\text{FlightInfo}, \text{WeatherAtAirports}, \text{Trajectory})$$

FlightInfo: `challenge_set.csv` and `final_submission_set.csv`

- Basic variables from `.csv` files but we did not use the callsign.
- Added variables:
  - Local time of departure/arrival computed from UTC time
  - Great circle distance, latitude/longitude from [ourairports.com]

WeatherAtAirport: METARs from [https://mesonet.agron.iastate.edu/ASOS/]

- Temperature and wind at departure and arrival airport
- Thunderstorms and fog at the arrival airport

Input Variables

$$\text{TOW} = \text{model}\left(\text{FlightInfo}, \text{WeatherAtAirports}, \text{Trajectory}\right)$$

Trajectory: OpenSky Network's ADS-B `2022-XX-XX.parquet` files

1. ADS-B Trajectory Filtering & Smoothing
2. Features Engineering:

Input Variables

$$\mathrm{TOW} = \mathrm{model}\left(\mathrm{FlightInfo}, \mathrm{WeatherAtAirports}, \mathrm{Trajectory}\right)$$

Trajectory: OpenSky Network's ADS-B `2022-XX-XX.parquet` files

1. ADS-B Trajectory Filtering & Smoothing
2. Features Engineering:
   - Wind along trajectory:
     Average value of the wind projected onto the ground speed

$$\mathrm{dot}\left(\vec{wind}, \vec{groundspeed}\right) / \|\vec{groundspeed}\|$$

Input Variables

$$\mathrm{TOW} = \mathrm{model}\left(\mathrm{FlightInfo}, \mathrm{WeatherAtAirports}, \mathrm{Trajectory}\right)$$

Trajectory: OpenSky Network's ADS-B `2022-XX-XX.parquet` files

1. ADS-B Trajectory Filtering & Smoothing
2. Features Engineering:
   - Wind along trajectory:
     Average value of the wind projected onto the ground speed

     $$\mathrm{dot}\left(\vec{wind}, \vec{groundspeed}\right) / \|\vec{groundspeed}\|$$

   - Climb phase:
     - $\mathrm{energyrate} = \frac{denergy}{dt}$, with $\mathrm{energy} = g_0 \mathrm{Altitude} + \frac{1}{2}\mathrm{TrueAirSpeed}^2$

Input Variables

$$\text{TOW} = \text{model}\left(\text{FlightInfo}, \text{WeatherAtAirports}, \text{Trajectory}\right)$$

Trajectory: OpenSky Network's ADS-B `2022-XX-XX.parquet` files

1. ADS-B Trajectory Filtering & Smoothing
2. Features Engineering:
   - Wind along trajectory:
     Average value of the wind projected onto the ground speed

$$\text{dot}\left(\vec{wind}, \vec{groundspeed}\right) / \|\vec{groundspeed}\|$$

   - Climb phase:
     - $\text{energyrate} = \frac{denergy}{dt}$, with $\text{energy} = g_0\text{Altitude} + \frac{1}{2}\text{TrueAirSpeed}^2$
     - Estimated mass using: $(\text{Thrust} - \text{Drag})\, V_a/\text{mass} = \text{energyrate}$
       Thrust and Drag model from OpenAP [Sun et al., 2020]
       Solving this equation $\Rightarrow$ roots of a 2nd degree polynomial [Alligier et al., 2013]

Input Variables

$$\text{TOW} = \text{model}\left(\text{FlightInfo}, \text{WeatherAtAirports}, \text{Trajectory}\right)$$

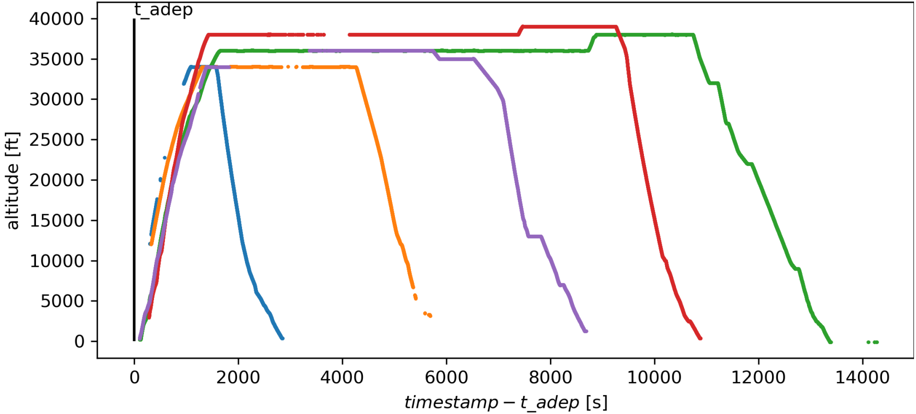
Trajectory: OpenSky Network's ADS-B `2022-XX-XX.parquet` files

1. ADS-B Trajectory Filtering & Smoothing
2. Features Engineering:
   - Wind along trajectory:
     Average value of the wind projected onto the ground speed

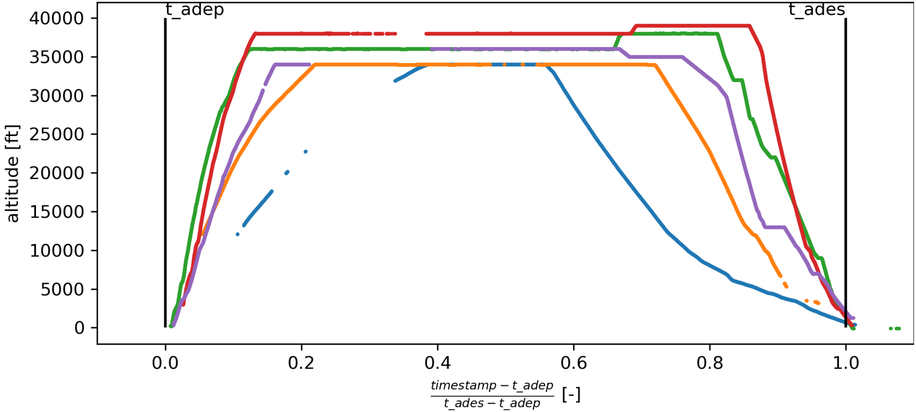
     $$\text{dot}\left(\vec{wind}, \vec{groundspeed}\right) / \|\vec{groundspeed}\|$$

   - Climb phase:
     - $\text{energyrate} = \frac{d\text{energy}}{dt}$, with $\text{energy} = g_0 \text{Altitude} + \frac{1}{2}\text{TrueAirSpeed}^2$
     - Estimated mass using: $(\text{Thrust} - \text{Drag})\, V_a / \text{mass} = \text{energyrate}$
       Thrust and Drag model from OpenAP [Sun et al., 2020]
       Solving this equation ⇒ roots of a 2nd degree polynomial [Alligier et al., 2013]
   - Flight profile: Cruise altitude and speed

Trajectory Features: Flight Profile

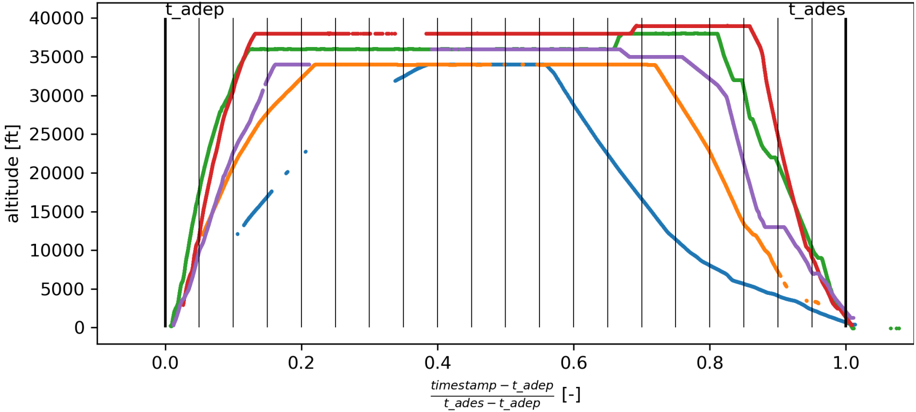


Trajectory Features: Flight Profile



Flight duration $t\_ades - t\_adep$

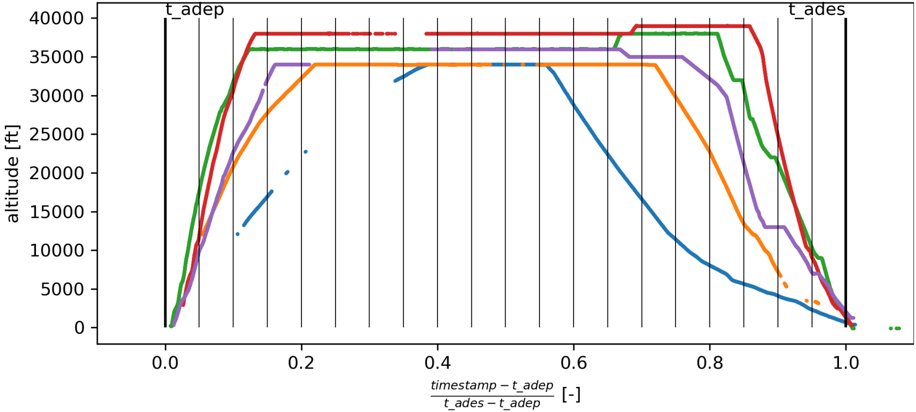
Trajectory Features: Flight Profile



Flight duration $t\_ades - t\_adep$, and 20 scaled temporal slices along the trajectory, starting from [0,5%] to [95%,100%]

- Cardinal (slice)
- $\text{median}_{i \in \text{slice}}$ Mach_i
- $\text{median}_{i \in \text{slice}}$ altitude_i
- $\text{altitude}_{\text{last(slice)}} - \text{altitude}_{\text{first(slice)}}$

Trajectory Features: Flight Profile

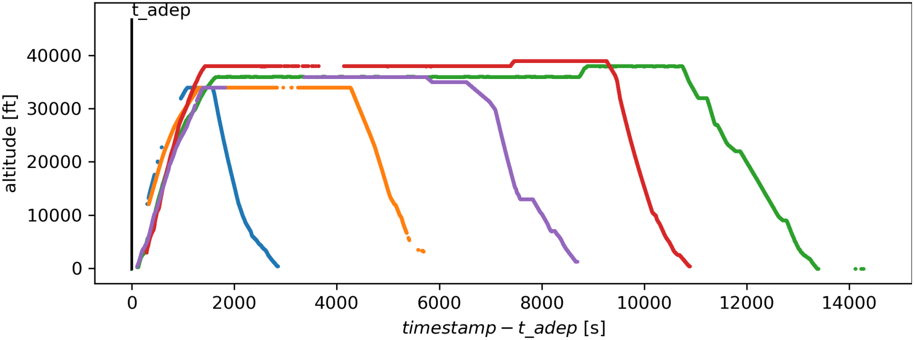


Flight duration $t\_ades - t\_adep$, and 20 scaled temporal slices along the trajectory, starting from [0,5%] to [95%,100%]

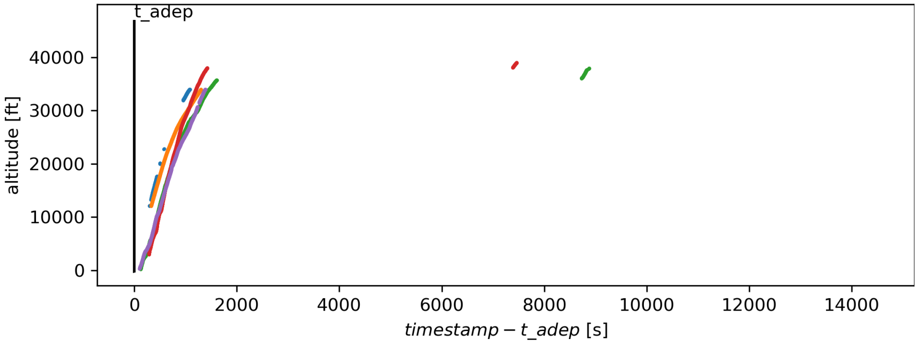
- Cardinal (slice)
- $\mathrm{median}_{i \in \mathrm{slice}}$ Mach_i
- $\mathrm{median}_{i \in \mathrm{slice}}$ $\mathrm{altitude}_i$
- $\mathrm{altitude}_{\mathrm{last(slice)}} - \mathrm{altitude}_{\mathrm{first(slice)}}$

This process generates $1 + 4 \times 20$ features

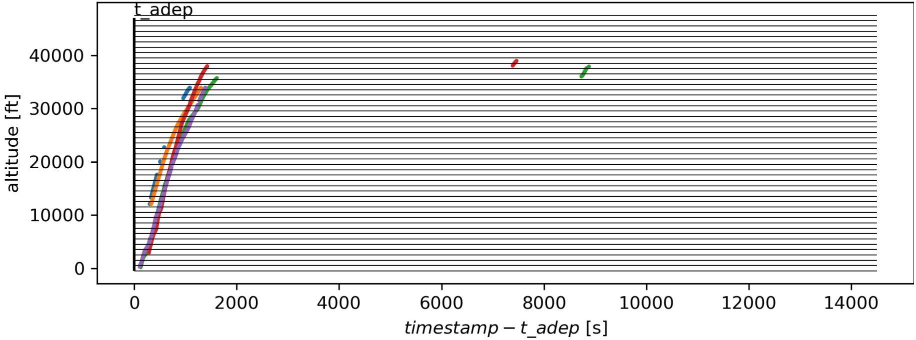
Trajectory Features: Climb Phase



Trajectory Features: Climb Phase



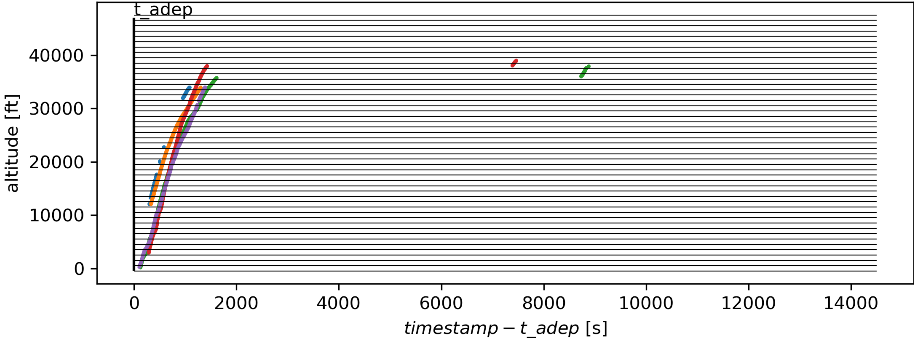
Trajectory Features: Climb Phase



48 vertical slices starting from [-500ft,500ft] to [46500ft,47500ft]

- Cardinal (slice)
- $\text{median}_{\mathrm{i}\in\mathrm{slice}}\ \Delta T_i$
- $\text{median}_{\mathrm{i}\in\mathrm{slice}}\ \text{TrueAirSpeed}_i$
- $\text{median}_{\mathrm{i}\in\mathrm{slice}}\ \text{ROCD}_i$
- $max_j ROCD_j - min_i ROCD_i$
- $\text{median}_{\mathrm{i}\in\mathrm{slice}}\ \text{mass}_i$
- $\min_{\mathrm{i}\in\mathrm{slice}}\ \text{energyrate}_i$
- $\text{median}_{\mathrm{i}\in\mathrm{slice}}\ \text{energyrate}_i$
- $\max_{\mathrm{i}\in\mathrm{slice}}\ \text{energyrate}_i$
- $\min_{\mathrm{i}\in\mathrm{slice}}\ \text{timestamp}_i - t\_adep$
- $t\_ades - \max_{\mathrm{i}\in\mathrm{slice}}\ \text{timestamp}_i$

Trajectory Features: Climb Phase



48 vertical slices starting from [-500ft,500ft] to [46500ft,47500ft]

- Cardinal (slice)
- $\mathrm{median}_{\mathrm{i}\in\mathrm{slice}}\ \Delta T_i$
- $\mathrm{median}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{TrueAirSpeed}_i$
- $\mathrm{median}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{ROCD}_i$
- $max_j ROCD_j - min_i ROCD_i$
- $\mathrm{median}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{mass}_i$
- $\mathrm{min}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{energyrate}_i$
- $\mathrm{median}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{energyrate}_i$
- $\mathrm{max}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{energyrate}_i$
- $\mathrm{min}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{timestamp}_i - t\_adep$
- $t\_ades - \mathrm{max}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{timestamp}_i$

$\Rightarrow$ Generates 11×48 features

Results

Using all these 611 input variables, we have an RMSE of 1,611 kg

Improved results through averaging models different random seeds:

- 10 Models (RMSE: 1,564 kg)
- 20 Models (RMSE: 1,561 kg)

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ?
- Cruise variables ?
- General climb variables (ROCD,etc) ?
- Mass estimates ?
- Energy rate variables ?

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ?
- Cruise variables ?
- General climb variables (ROCD,etc) ?
- Mass estimates ?
- Energy rate variables ?

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
×	×	×	×	×	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ?
- Cruise variables ?
- General climb variables (ROCD,etc) ?
- Mass estimates ?
- Energy rate variables ?

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
$\times$	$\times$	$\times$	$\times$	$\times$	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ?
- Cruise variables ?
- General climb variables (ROCD,etc) ?
- Mass estimates ?
- Energy rate variables ?

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✗	✗	✗	✗	✗	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ? No !
- Cruise variables ?
- General climb variables (ROCD,etc) ?
- Mass estimates ?
- Energy rate variables ?

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
×	×	×	×	×	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ? No !
- Cruise variables ? Not that much
- General climb variables (ROCD,etc) ?
- Mass estimates ?
- Energy rate variables ?

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
×	×	×	×	×	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ? No !
- Cruise variables ? Not that much
- General climb variables (ROCD,etc) ? Yes, somewhat
- Mass estimates ?
- Energy rate variables ?

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
×	×	×	×	×	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ? No !
- Cruise variables ? Not that much
- General climb variables (ROCD,etc) ? Yes, somewhat
- Mass estimates ? Yes, somewhat
- Energy rate variables ?

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✗	✗	✗	✗	✗	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog variables ? No !
- Cruise variables ? Not that much
- General climb variables (ROCD,etc) ? Yes, somewhat
- Mass estimates ? Yes, somewhat
- Energy rate variables ? Yes !!

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✗	✗	✗	✗	✗	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686
✓	✓	✓	✓	✓	1611

Conclusion

Predicting TOW with a good accuracy is possible

What Worked ?

- Information extracted from climb phase (energy rate !)
- Filtering and smoothing (??) + many slices (??)

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✓	✓	✓	✗	✗	1954

Conclusion

Predicting TOW with a good accuracy is possible

What Worked ?

- Information extracted from climb phase (energy rate !)
- Filtering and smoothing (??) + many slices (??)

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✓	✓	✓	✗	✗	1954

Features Not so Useful in our Solution ?

- Thunder and fog at arrival airport
- Cruise features

Conclusion

Predicting TOW with a good accuracy is possible

What Worked ?

- Information extracted from climb phase (energy rate !)
- Filtering and smoothing (??) + many slices (??)

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✓	✓	✓	✗	✗	1954

Features Not so Useful in our Solution ?

- Thunder and fog at arrival airport
- Cruise features

Perspective ?

- Is it possible to extract more info from cruise and descent phases ?
- A benchmark that will be used by researchers on future study ?!

Conclusion

Predicting TOW with a good accuracy is possible

What Worked ?

- Information extracted from climb phase (energy rate !)
- Filtering and smoothing (??) + many slices (??)

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✓	✓	✓	✗	✗	1954

Features Not so Useful in our Solution ?

- Thunder and fog at arrival airport
- Cruise features

Perspective ?

- Is it possible to extract more info from cruise and descent phases ?
- A benchmark that will be used by researchers on future study ?!

Thanks to the organizers for this nice data challenge, it has been fun ! :-)

Thank you for your attention

[Alligier et al., 2013]
Ground-based estimation of aircraft mass, adaptive vs. least squares method.

[Sun et al., 2020]
Openap: An open-source aircraft performance model for air transportation studies and simulations.

Trajectory features

Climbing phase

48 vertical slices starting from [-500ft,500ft] to [46500ft,47500ft]
For each slice:

- $\mathrm{Cardinal}\,(\mathrm{slice})$ - number of points in the slice
- $\mathrm{median}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{ROCD}_i$
- $max_j ROCD_j - min_i ROCD_i$
- $\mathrm{median}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{TrueAirSpeed}_i$
- $\mathrm{median}_{\mathrm{i} \in \mathrm{slice}}\ \Delta T_i$
- $\mathrm{min}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{energyrate}_i$
- $\mathrm{median}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{energyrate}_i$
- $\mathrm{max}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{energyrate}_i$
- $\mathrm{min}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{timestamp}_i - t\_adep$
- $t\_ades - \mathrm{max}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{timestamp}_i$
- $\mathrm{median}_{\mathrm{i} \in \mathrm{slice}}\ \mathrm{mass}_i$

Trajectory features

Flight profile

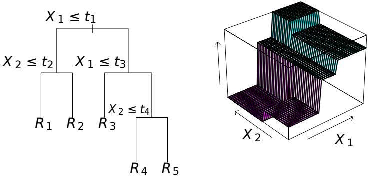
20 scaled temporal slices along the trajectory, starting from [0,5%] to [95%,100%]
For each slice:

- $t\_ades - t\_adep$ - the scaling factor and flight duration
- $\mathrm{Cardinal}\,(\mathrm{slice})$ - number of points in the slice
- $\mathrm{median}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{Mach}_i$
- $\mathrm{median}_{\mathrm{i}\in\mathrm{slice}}\ \mathrm{altitude}_i$
- $\mathrm{altitude}_{last(slice)} - \mathrm{altitude}\_first(slice)$

Machine Learning Model

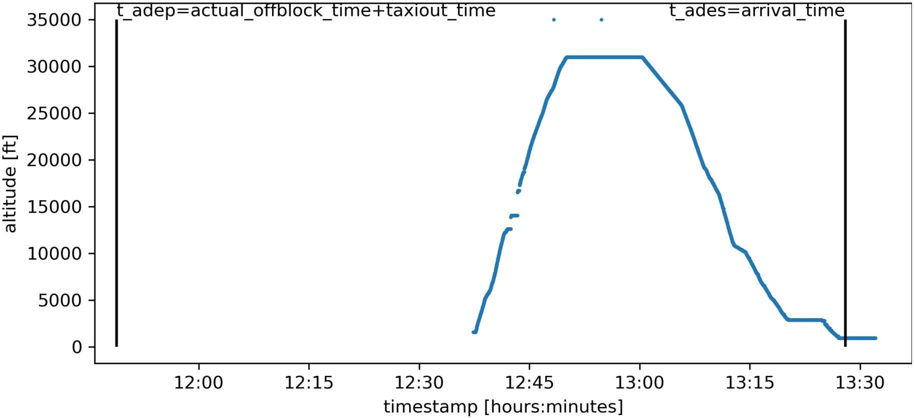
- Theoretical framework: stochastic gradient boosting
  [1]
- Gradient-boosted regression trees:
  - Sum of weak prediction models $h_m(\mathrm{x}) = h_{m-1}(\mathrm{x}) + \nu\, t_m(\mathrm{x})$
  - with $t_m(x) = \sum_{R_j \in T_m} \gamma_{mj} \mathbb{1}_{R_j}(\mathrm{x})$ a small tree

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

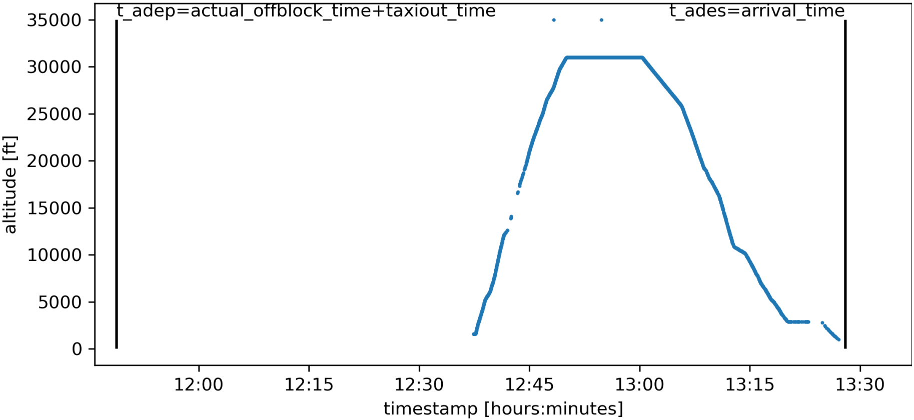


  - Iterative training: small tree $t_m(\mathrm{x})$ tuned on residuals of previous model h_{m-1}, with random sampling

ADS-B Trajectory Filtering & Smoothing

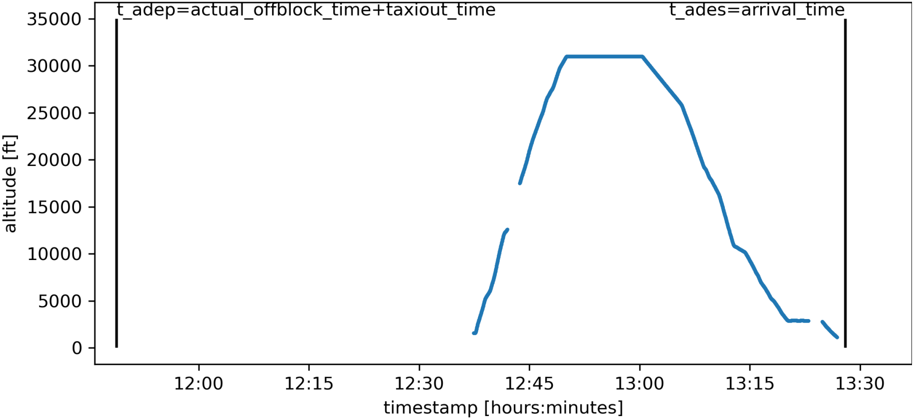


ADS-B Trajectory Filtering & Smoothing



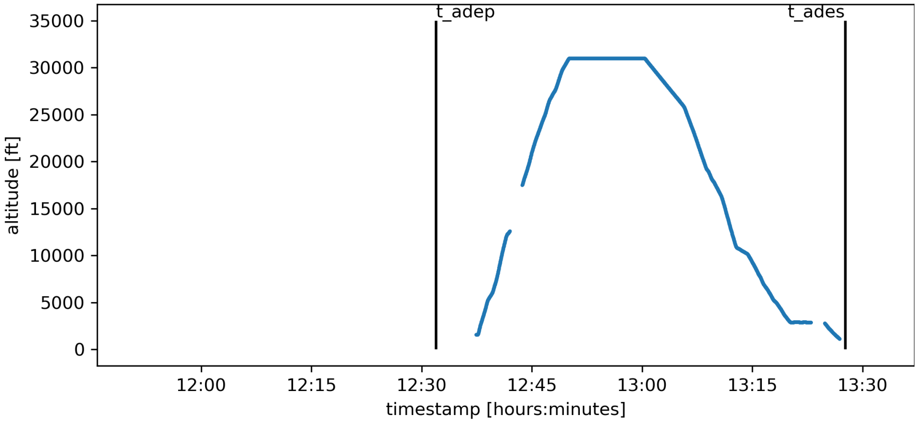
1. Filtering out repeated measurements

ADS-B Trajectory Filtering & Smoothing



1. Filtering out repeated measurements
2. Filtering out measurements associated with a second order derivative above a threshold
3. Trajectories are smoothed using cubic splines (csaps library)

ADS-B Trajectory Filtering & Smoothing



1. Filtering out repeated measurements
2. Filtering out measurements associated with a second order derivative above a threshold
3. Trajectories are smoothed using cubic splines (csaps library)
4. Correct take-off/landing datetimes

Follow-up: Ablation Study of the Built Features

Which features do the heavy lifting ?

- Thunderstorm and fog, No !
- Do cruise variables are that useful ? not that much
- Do mass estimates are that useful ? Yes, somewhat
- Do energy rate variables are that useful ? Yes !!

TS & Fog	Cruise	Climb			RMSE [kg]
		Other	Mass	energy_rate	
✓	✓	✓	✓	✓	1611
✗					1606
	✗				1610
			✗		1609
				✗	1721
✗	✗	✗	✗	✗	3147
✓					3147
	✓				2489
		✓			1978
			✓		1936
				✓	1686