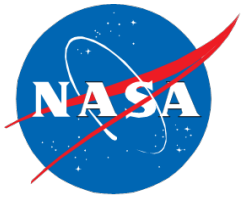


ATD-2 Integrated Arrival/ Departure/Surface (IADS) System

Machine Learning Services

Impeded Taxi Out Time Prediction Model (ITOM)

Unimpeded Taxi Out Time Prediction Model (UTOM)

Alexandre Amblard, Sarah Youlton

Universities Space Research Association (USRA) - NASA Academic Mission Services (NAMS)

NASA Ames Research Center, Moffett Field, CA

William J. Coupe

NASA Ames Research Center, Moffett Field, CA

August 2021

Table of Contents

Model Name

Two Taxi Out models are summarized in this document:

- Impeded Taxi Out Time Prediction (ITOM)
- Unimpeded Taxi Out Time Prediction (UTOM)

Problem Statement

Impeded Taxi Out

The Impeded Taxi Out model predicts the transit time of aircraft on the surface from pushback to spot crossing and take-off while considering surface traffic. The following sub-models compute this transit time:

- Impeded Ramp Taxi Out time: predicts the impeded taxi time between the gate and the spot crossing
- Impeded Active Movement Area (AMA) Taxi Out time: predicts the impeded taxi time between the spot crossing and the runway
- Impeded Full Taxi Out time: predicts the impeded taxi time between the gate and the runway

Unimpeded Taxi Out

The Unimpeded Taxi Out time prediction service is predicting the transit time of aircraft while unimpeded on the surface from pushback to spot crossing and take-off. Unimpeded Taxi Out Time estimates are essential quantities for a scheduler to build an efficient schedule, allowing aircraft with lower unimpeded taxi-out time (i.e. that could reach the runway faster) to be put first.

- Unimpeded Ramp Taxi Out time: predicts the unimpeded taxi time between the gate and the spot crossing
- Unimpeded AMA Taxi Out time: predicts the unimpeded taxi time between the spot crossing and the runway
- Unimpeded Full Taxi Out time: predicts the unimpeded taxi time between the gate and the runway

Technical Approach

The prediction services are contained in a pickled scikit-learn pipeline that applies some feature engineering and the model fit. The pickle files, containing the fitted models that will be deployed for prediction, are produced with kedro pipelines that prepare the training and testing data before being fed to the scikit-learn pipeline.

Impeded Taxi Out

The current feature engineering step is one-hot encoding **departure_runway_actual** and **carrier** features. The **departure_stand_actual** feature is encoded with a clustering technique that groups gates in C clusters depending on the median target taxi-time for N runways. The current clustering algorithm used is the Ward hierarchical clustering from scikit-learn library (AgglomerativeClustering) with N and C set to 4 and 8, respectively. The **departure_stand_airline_time** is placed into ten minute bins based on hour and minute of the datetime.

The engineered features are fed to a Gradient Boosted Tree regression algorithm, specifically the XGBoost library implementation (XGBRegressor) for the Full, AMA, and Ramp Taxi Out time prediction. Non-ASDEX (Airport Surface Detection Equipment, Model-X) airports (such as KDAL) have no surface data and only the full taxi model is fitted, the AMA taxi time model is a copy of the full taxi model and the ramp taxi time model returns 0.

Unimpeded Taxi Out

One of the key steps of the kedro training harness is to filter unimpeded flights for unimpeded AMA and full taxi using ASDEX information about aircraft ground speed. The filter requirement is that the ground speed stays above a 4 knot threshold 90% of the taxi duration. Since ramp ground speed is unavailable, this filter is not applied in the ramp taxi time kedro pipeline, instead a filter based on an estimate of the surface congestion (aircraft surface count) is applied to the data. Non-ASDEX airports (such as KDAL) have no surface data and only the full taxi is fitted, the AMA taxi time model is a copy of the full taxi model and the ramp taxi time model returns 0. For the full taxi time of non-ASDEX airports, the kedro pipeline applies a filter to select taxi-time between the 10th and 30th percentile in order to select flights most likely to be unimpeded.

The current feature engineering step is one-hot encoding **departure_runway_actual** and **carrier** features. The **departure_stand_actual** feature is encoded with a clustering technique that groups gates in C clusters depending on the median target taxi-time for N runways. The current clustering algorithm used is the Ward hierarchical clustering from scikit-learn library (AgglomerativeClustering) with N set to 3 and C varying from 3 to 11, depending on the airport.

The engineered features are fed to a Gradient Boosted Tree regression algorithm, specifically the XGBoost library implementation (XGBRegressor) for the all Taxi Out time prediction.

Model Features

Impeded Taxi Out

Feature	Description	Sample Value
departure_runway_actual	Departure runway ID	17L
departure_stand_actual	Gate ID	B12
carrier	Carrier (airline) ID	AAL
total_arrivals_on_surface	Number of arrival flights on the surface	10
total_departures_on_surface	Number of departure flights on the surface	10
departure_stand_airline_time	Itime	2020-08-01 00:00:00

Unimpeded Taxi Out

Feature	Description	Sample Value
departure_runway_actual	Departure runway ID	17L
departure_stand_actual	Gate ID	B12
carrier	Carrier (airline) ID	AAL

Model Inputs & Outputs

See OpenAPI specification in the appendix.

Data Sets

For both Impeded and Unimpeded Taxi Out models, input data are queried from several databases. **departure_runway_actual** is queried from the matm_flight_summary (fuser on tbd-warehouse) and runways (tbd_warehouse_runways on int7) tables, the runways table values taking precedence over matm_flight_summary values. **departure_stand_actual** and **carrier** are queried from the matm_flight_summary table. Furthermore for matm_flight_summary values, **departure_runway_actual** and **departure_stand_actual** are further defined by the first non-null value of the following columns: departure_runway_actual, departure_runway_user, departure_runway_assigned, departure_runway_airline and departure_stand_actual, departure_stand_user, departure_stand_airline for runway and stand respectively.

For the Impeded Taxi Out models, **total_arrivals_on_surface** and **total_departures_on_surface** are computed at pushback (as indicated by departure_stand_actual_time from matm_flight_summary table) using departure/arrival_*[stand/runway/movement_area]*_actual_time from matm_flight_summary and incrementing and decrementing the aircraft count as aircraft enter/exit an area (ramp or AMA). This algorithm can potentially be a problem if the surface is not empty at the beginning of the count. **departure_stand_airline_time** is queried from the matm_flight_summary table.

For both Impeded and Unimpeded Taxi Out models, the target values for the ramp, AMA and full taxi time pipeline are **actual_departure_ramp_taxi_time**, **actual_departure_ama_taxi_time**, **actual_departure_full_taxi_time**. These quantities are derived by subtracting two of the following quantities: departure_stand_actual_time, departure_movement_area_actual_time, departure_runway_actual_time. All these timestamps are queried from the matm_flight_summary table and departure_runway_actual_time is also queried from the runways table which values take priority over the matm_flight_summary ones.

The training/validation/test data are queried from June 1st 2019 08:00 (UTC) to December 30th 08:00 (UTC) for KDFW and KCLT, and from August 1st 2020 08:00 (UTC) to December 31st 2020 08:00 (UTC) for KDAL. 80% of the data are used for training/validation, and 20% are used for testing.

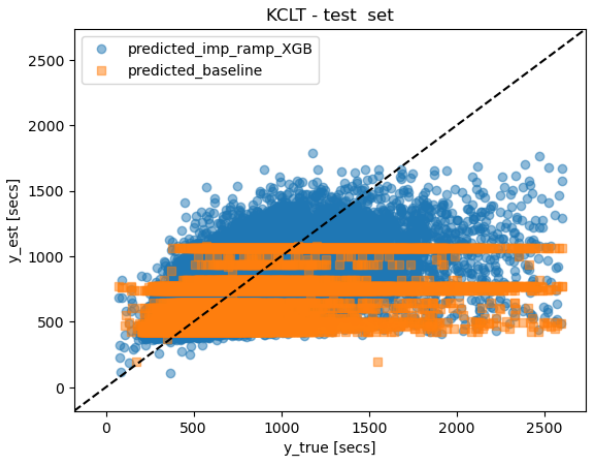
Model Results / Evaluation

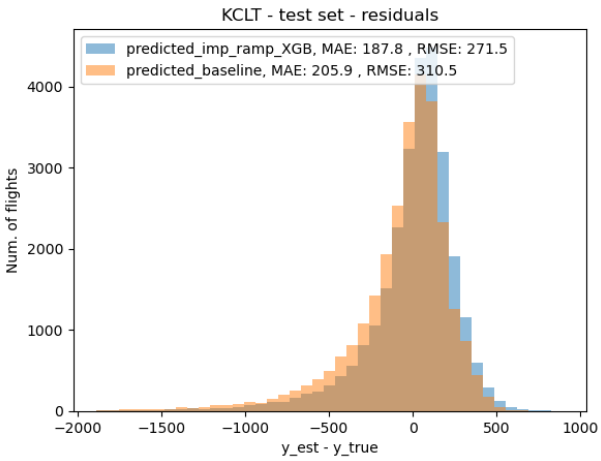
Impeded Taxi Out

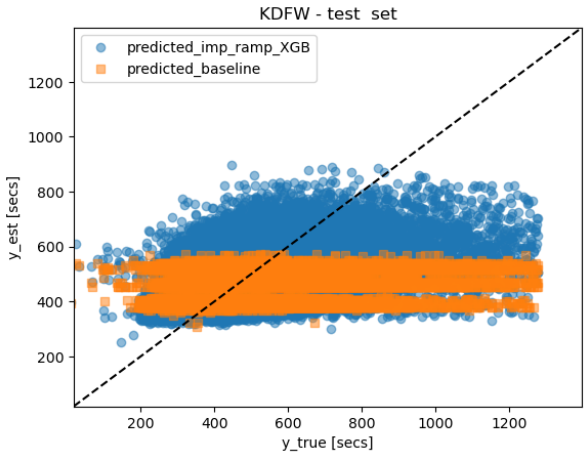
In the following table, MAD stands for Median Absolute Deviation, and it is multiplied by 1.4826 to match the standard deviation value if the residual distribution is Gaussian. Residuals are calculated by subtracting estimated values from truth values (negative residuals indicate over-estimated taxi times). Truth values are actual taxi time. All quantities in the following tables are in seconds.

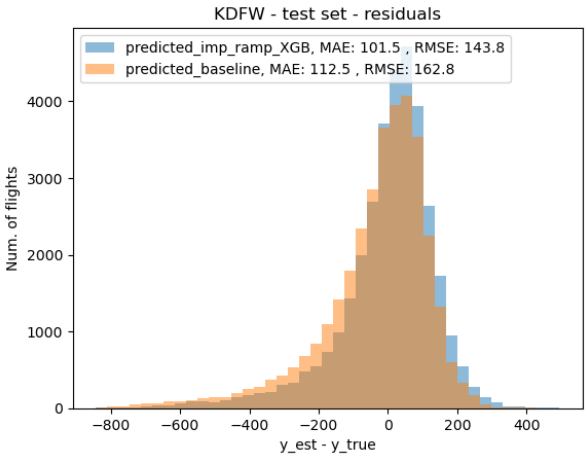
Ramp

Airport	MAD of Residual (test)	Median of Residual (test)	Median Taxi Time (test)
KCLT	178	-53	684
KDFW	101	-24	455
KDAL	-	-	-



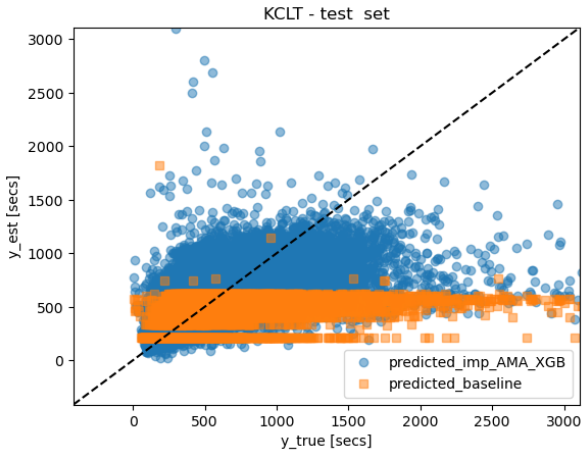


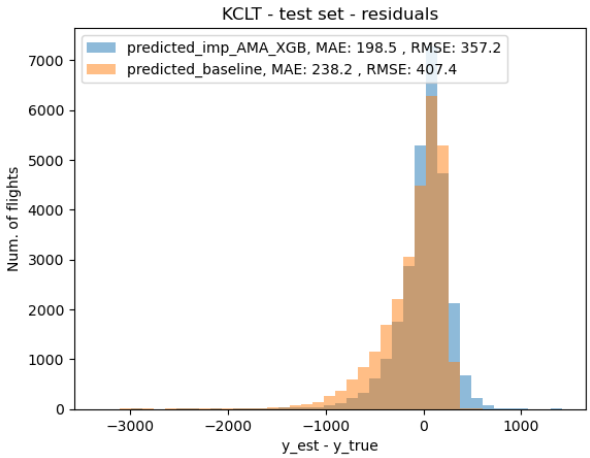


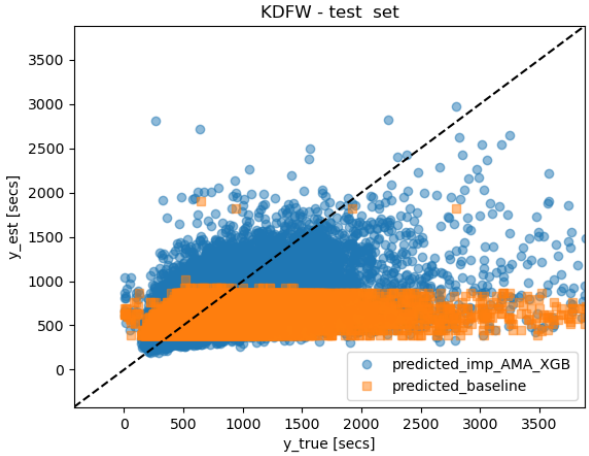


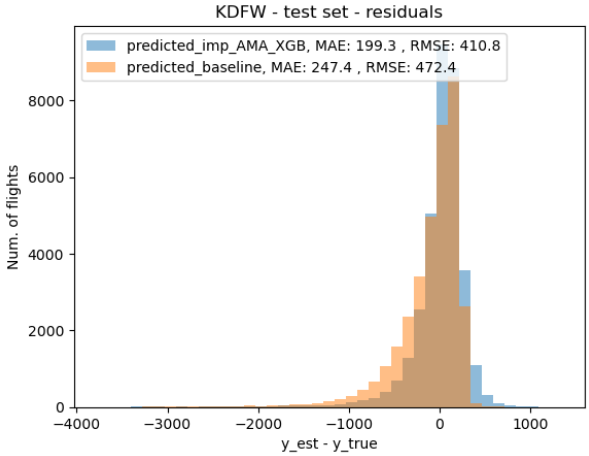
Active Movement Area

Airport	MAD of Residual (test)	Median of Residual (test)	Median Taxi Time (test)
KCLT	188	-49	581
KDFW	174	-56	665
KDAL (copy of full)	129	-40	638



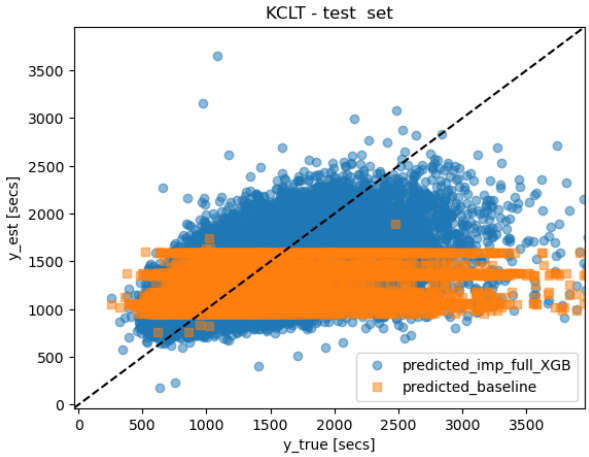


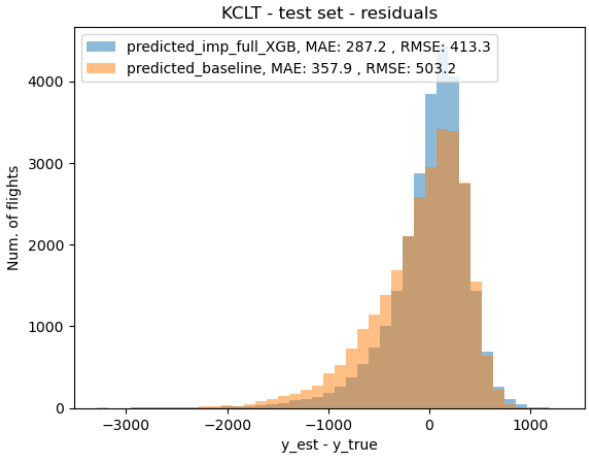


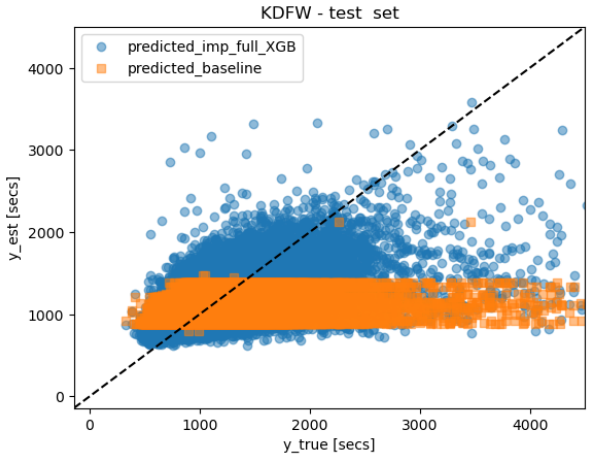


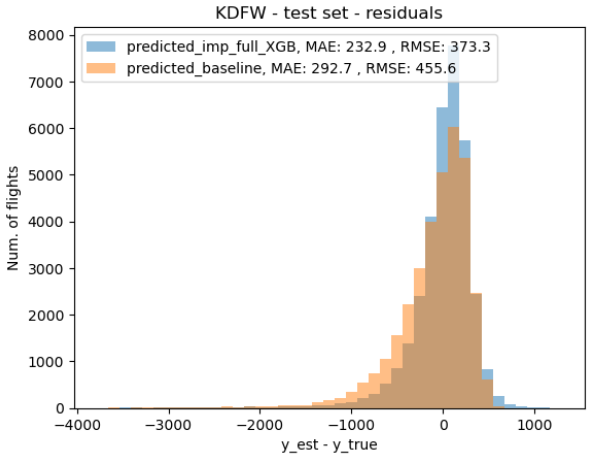
Full (Ramp + AMA)

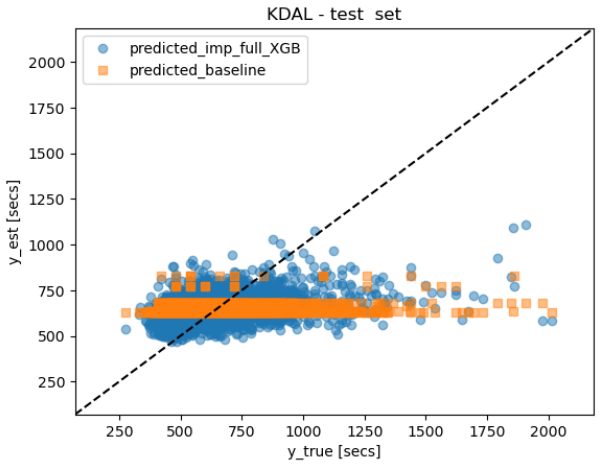
Airport	MAD of Residual (test)	Median of Residual (test)	Median Taxi Time (test)
KCLT	297	-72	1274
KDFW	224	-61	1119
KDAL	129	-40	638

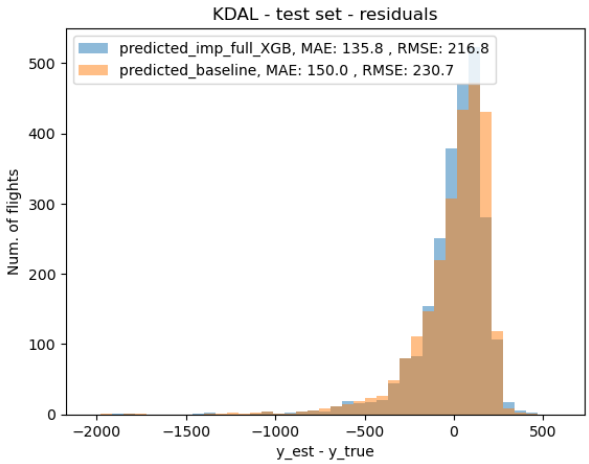










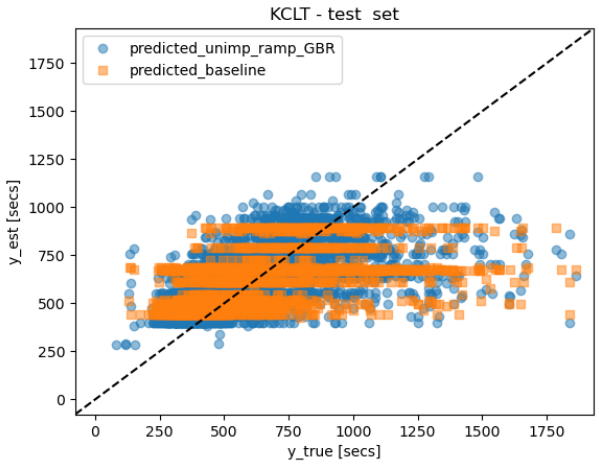


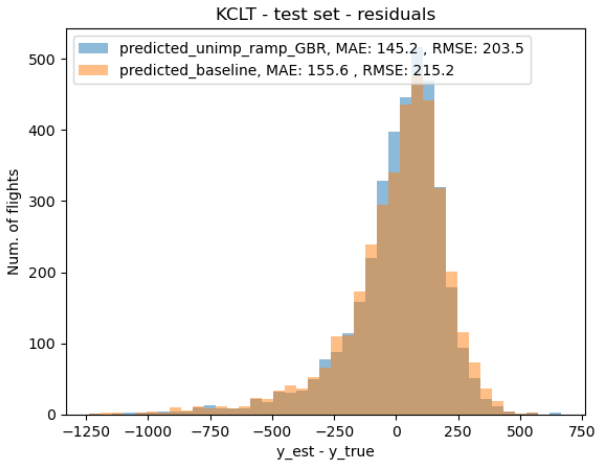
Unimpeded Taxi Out

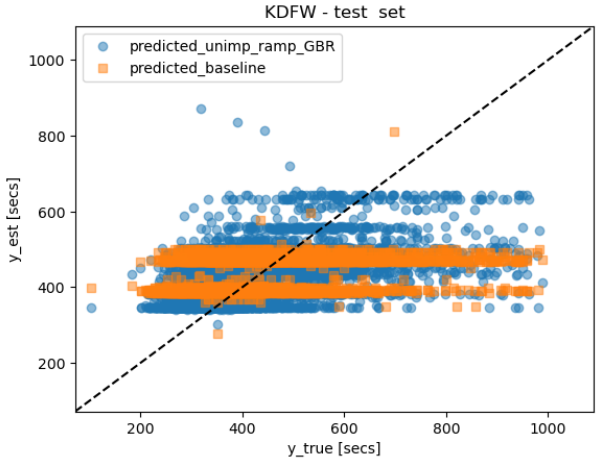
In the following table, MAD stands for Median Absolute Deviation, and it is multiplied by 1.4826 to match the standard deviation value if the residual distribution is Gaussian. Residuals are calculated by subtracting estimated values from truth values (negative residuals indicate over-estimated taxi times). Truth values are actual taxi time in the first columns, whereas they are Surface Trajectory Based Operations (STBO) unimpeded taxi time in the latter columns (STBO residual). Both of these truth values are imperfect proxies to the true unimpeded taxi times. All quantities in the following tables are in seconds.

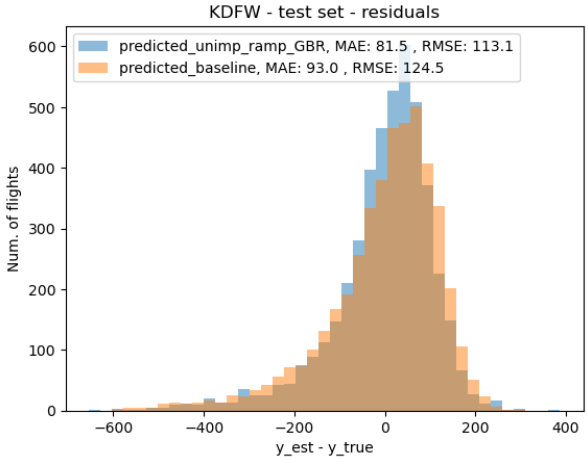
Ramp

Airport	MAD of Residual (test)	Median of Residual (test)	MAD of STBO Residual (test)	Median of STBO Residual (test)	Median Taxi Time (test)
KCLT	144	-41	100	-67	635
KDFW	83	-18	79	8	426
KDAL	-	-	-	-	-



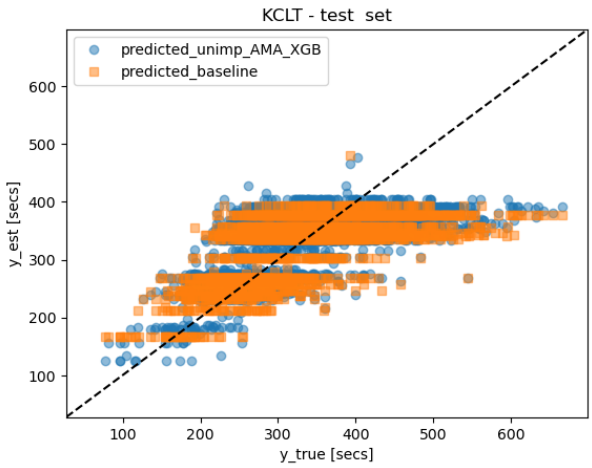


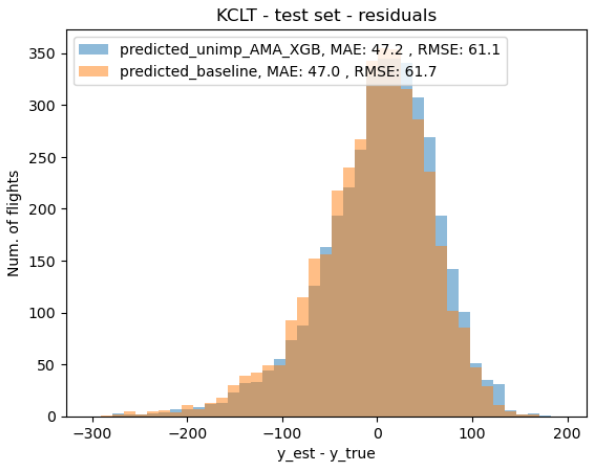


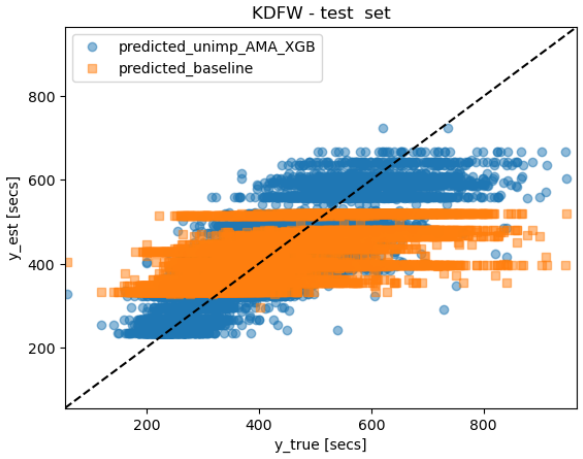


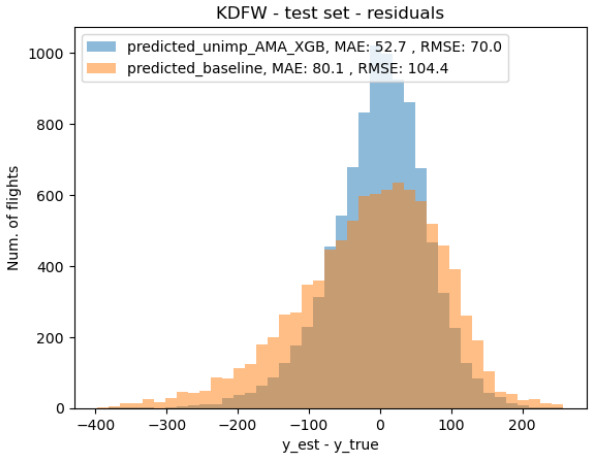
Active Movement Area

Airport	MAD of Residual (test)	Median of Residual (test)	MAD of STBO Residual (test)	Median of STBO Residual (test)	Median Taxi Time (test)
KCLT	55	-9	21	-42	332
KDFW	62	-3	39	-69	404
KDAL	-	-	-	-	-



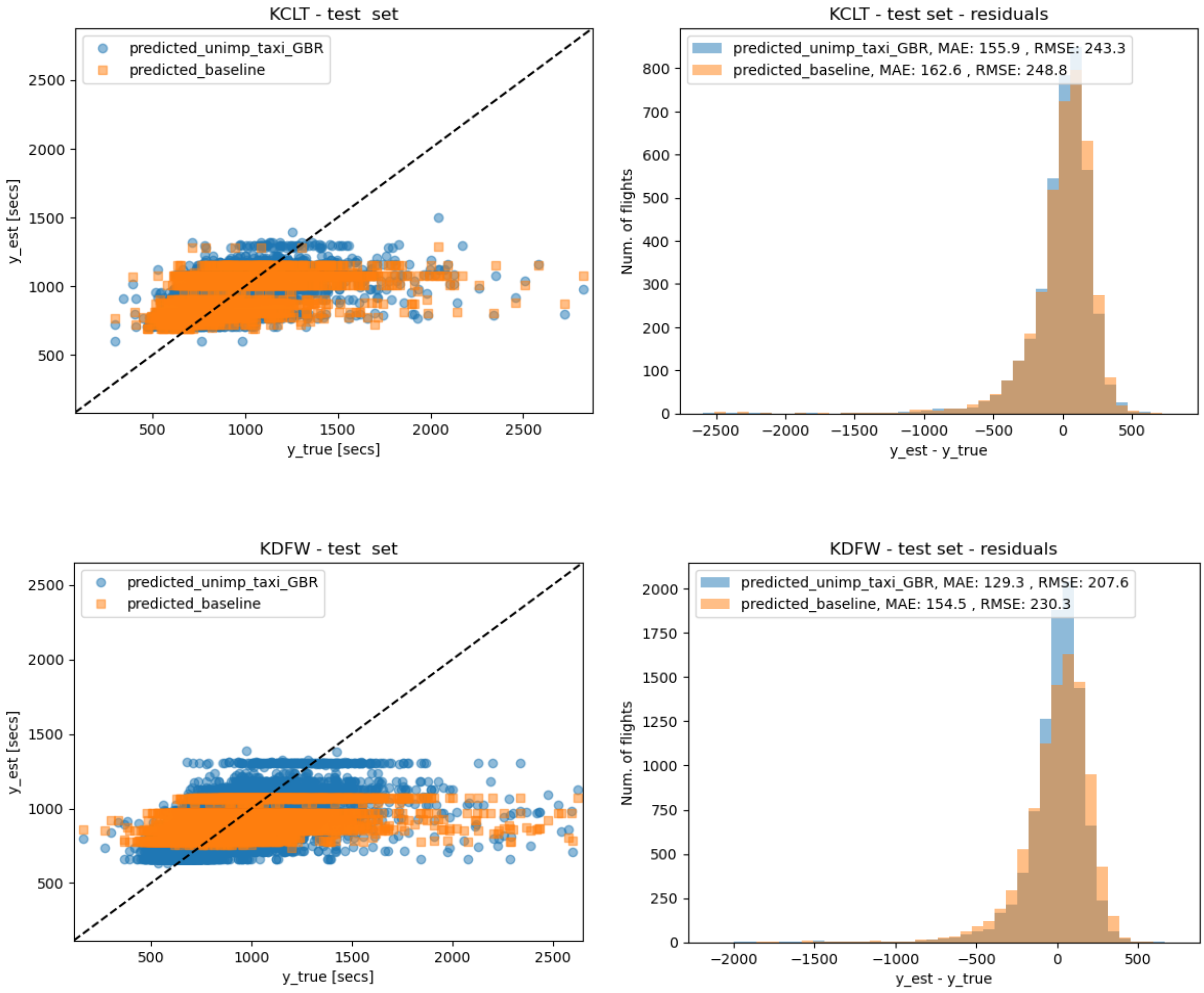






Full (Ramp + AMA)

Airport	MAD of Residual (test)	Median of Residual (test)	MAD of STBO Residual (test)	Median of STBO Residual (test)	Median Taxi Time (test)
KCLT	153	-41	79	-102	867
KDFW	129	-28	95	-79	826
KDAL	26	-3	-	-	489



Open Source Repository

https://github.com/nasa/ML-airport-taxi-out

Reference Documentation

Amblard, A., Youlton, S., Coupe, W., “Real-time Unimpeded Taxi Out Machine Learning Service,” AIAA AVIATION Forum, Washington, DC, USA, 2021.

Appendix: OpenAPI Specification

Explore

Taxi Out OVC Client

1.0.1-dev11 OAS3

Taxi Out OVC Client

Servers

http://localhost:9110/ - Out In OVC Client

Computed URL: `http://localhost:9110/`

Server variables

port 9110

default

POST /unimpeded/full/taxi-out

POST /unimpeded/ramp/taxi-out

POST /unimpeded/ama/taxi-out

Schemas

TaxiInResponseFlight {

error	string *example: ABC123.CLT.MEX* error description
gufi	string *example: ABC123.CLT.MEX* gufi
pred	number($float) *example: 304.0* taxi out prediction value

}

TaxiOutResponse {

airport*	string *example: KDFW* airport name
flights*	[...]

}

UnimpededTaxiInRequestFlight {

aircraftType	string *example: E170* *nullable: true* aircraft type
carrier	string *example: AAL* *nullable: true* carrier name
departureRunwayActual	string *example: 17C* *nullable: true* departure runway actual
departureStandActual	string *example: C17* *nullable: true* adeparture stand actual
gufi	string *example: ABC123.CLT.MEX* *nullable: true* gufi

}

UnimpededTaxiOutRequest {

airport*	string *example: KDFW* airport name	
flights*	[...]	

}