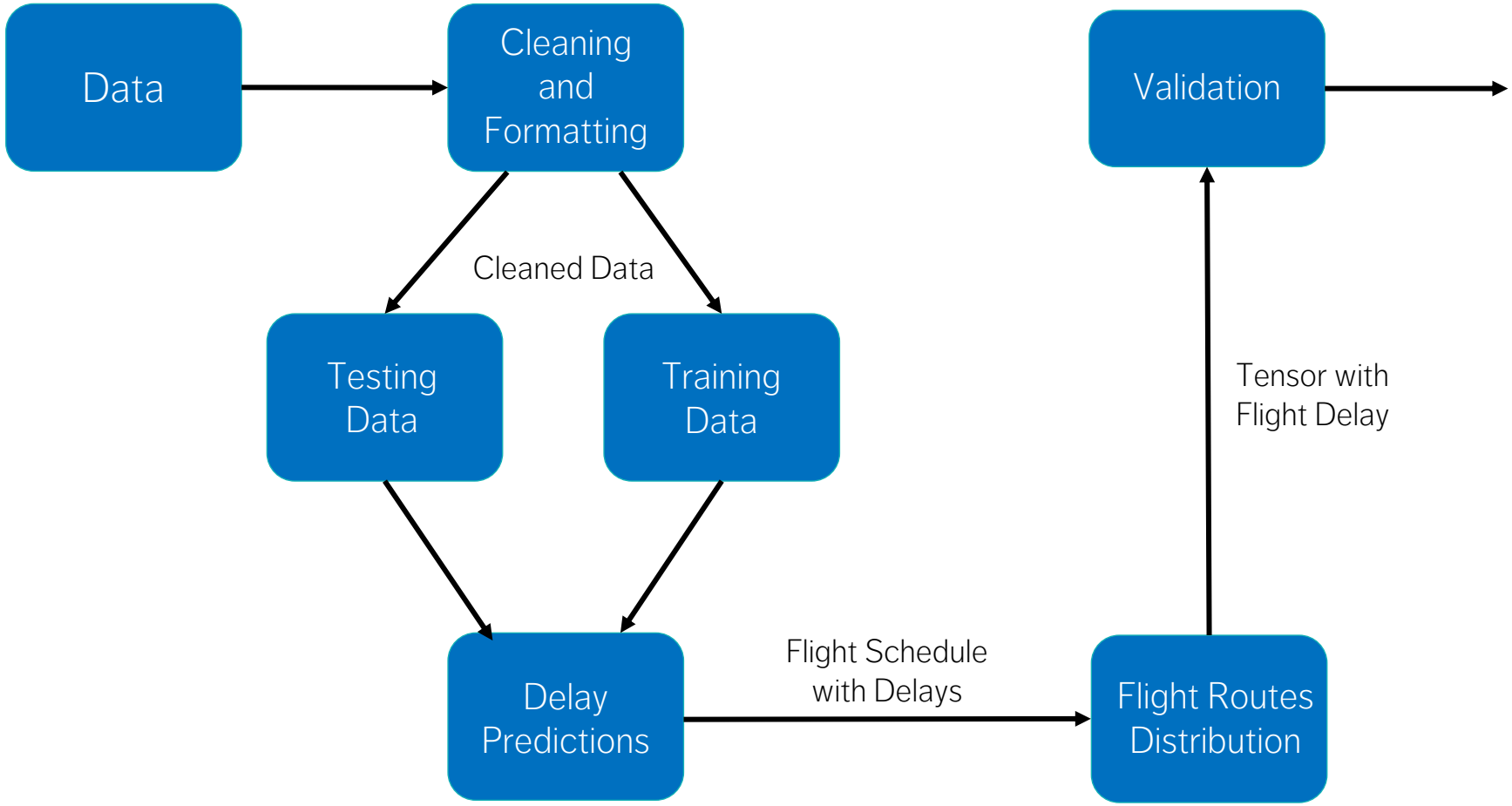
Flight Route Predictive Analytics Model

V. Kishore, S. Patel, D. Pramod, P. Boily
Idlewyld Analytics and Consulting Services

IDLEWYLD

Methodology

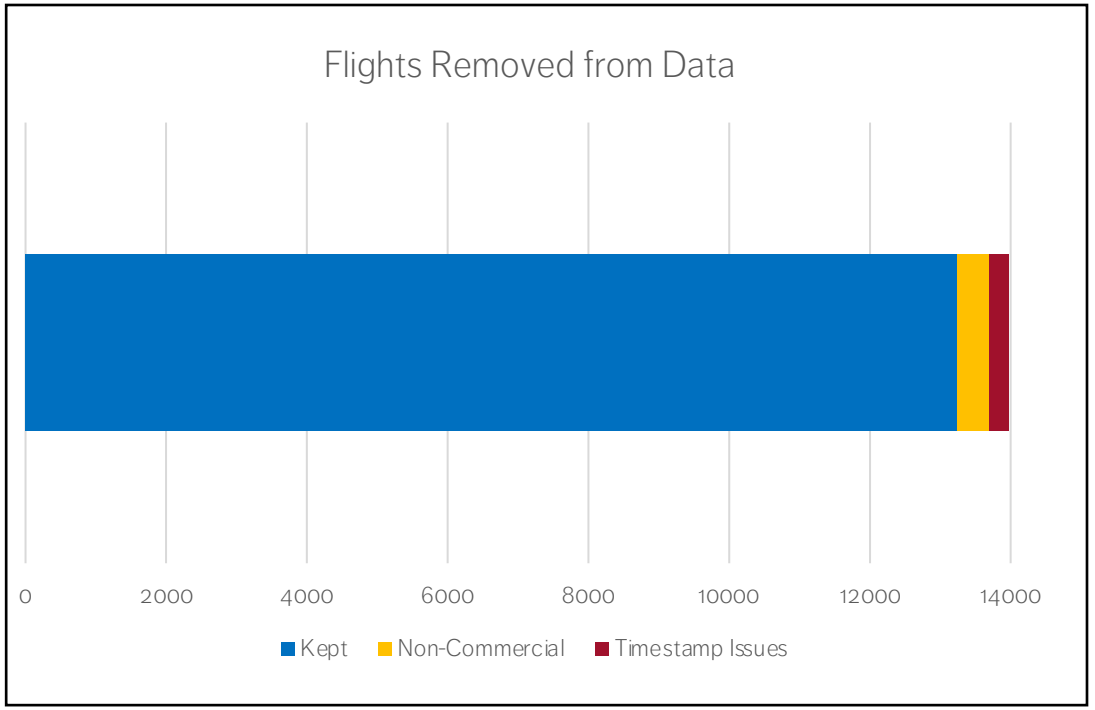


IDLEWYLD

Data Cleaning

Removed non-commercial flights as there was not enough historical data and did not pertain to project.

Also removed 269 flights from dataset with first timestamp occurring 15 min after departure.



IDLEWYLD

Formatting

Only certain variables kept including:

Flight ID	Filed Departure Time
Origin	Actual Departure Time
Destination	Date

Latitude and Longitude were replaced by index variables

- based on “flat” map split into $2° \times 2°$ regions

Time variables were replaced by time indices

- each index represents a 5-min interval

[Flightradar24.com and Telesat data]

Sample Data

	Flight ID	Ident	Type	Origin	Destinatio	Filed Departure Time (UTC)	Actual Departure	Filed Arrival Time (UTC	Actual Arrival Tim	Time (UTC)	Latitude	Longitude
2	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:29	12.0058	-61.749
3	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:30	12.0069	-61.698
4	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:30	12.0272	-61.672
5	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:31	12.046	-61.665
6	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:31	12.0808	-61.651
7	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:31	12.0969	-61.644
8	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:32	12.1587	-61.619
9	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:35	12.3609	-61.538
10	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:36	12.4587	-61.499
11	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:37	12.5553	-61.461
12	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:37	12.6094	-61.439
13	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:38	12.6575	-61.42
14	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:39	12.7612	-61.378
15	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:40	12.8656	-61.336
16	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:41	12.9727	-61.293
17	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:42	13.0827	-61.249
18	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:42	13.1379	-61.227
19	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:43	13.1925	-61.204
20	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:44	13.306	-61.157
21	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:44	13.363	-61.133
22	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:45	13.4218	-61.108
23	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:45	13.4782	-61.084
24	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:46	13.5388	-61.059
25	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:46	13.5972	-61.034
26	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:47	13.7169	-60.984
27	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:48	13.7756	-60.949
28	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:49	13.8869	-60.877
29	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:50	13.997	-60.806
30	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:51	14.1168	-60.728
31	9HVJY-15	VJT930	GLEX	TGPY	UUWW	7/9/2018 1:00	7/9/2018 1:28	7/9/2018 12:22	7/9/2018 12:50	7/9/2018 1:51	14.1725	-60.692

IDLEWYLD

Snapshot of Data Prior to Cleaning

Flight ID	Ident	Type	Origin	Destination	Filed Departure Time (UTC)	Actual Departure Time (UTC)	Filed Arrival Time (UTC)	Actual Arrival Time (UTC)	Time (UTC)	Latitude	Longitude	Groundspeed (kts)	Altitude (ft)	Rate	Course	Direction	Facility Name
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:29:13	12.00580025	-61.74890137	78	25000		89	East	TZP
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:30:09	12.00689983	-61.69779968	191	4200	-12848	86	East	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:30:45	12.02719975	-61.67219925	225	5300	2073	17	North	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:31:04	12.04599953	-61.6651001	221	6100	2333	25	Northeast	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:31:39	12.08080006	-61.65060043	240	7400	1882	22	Northeast	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:31:55	12.09689999	-61.64400101	260	7700	1545	22	Northeast	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:32:45	12.15869999	-61.61920166	293	9100	1641	22	Northeast	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:35:07	12.36089993	-61.53839874	365	12950	1626	21	North	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:36:07	12.45870018	-61.4992981	383	14575	1775	22	Northeast	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:37:07	12.55529976	-61.46070099	384	16500	1883	21	North	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:37:37	12.6093998	-61.43899918	388	17400	1625	21	North	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:38:07	12.65750027	-61.4197998	394	18125	1550	21	North	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:39:07	12.76119995	-61.37829971	399	19725	1500	21	North	FlightAware ADS-B
9HVJY-1531063700-dlad-16009470:6	VJT930	GLEX	TGPY	UUWW	2018-07-09 1:00	2018-07-09 1:28	2018-07-09 12:22	2018-07-09 12:50	07/09/2018 01:40:07	12.86559963	-61.33649826	409	21125	1375	21	North	FlightAware ADS-B

Snapshot of Data After Cleaning

Flight.ID	ts.index	lat.index	lon.index	Origin	Destination	Filed Departure Date	Filed Departure Time	Actual Departure Date	Actual Departure Time	Date	dep.index
A6DFR-1531579944-adhoc-0:0	1	33	27	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15
A6DFR-1531579944-adhoc-0:0	2	33	27	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15
A6DFR-1531579944-adhoc-0:0	3	33	28	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15
A6DFR-1531579944-adhoc-0:0	4	33	28	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15
A6DFR-1531579944-adhoc-0:0	5	34	28	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15
A6DFR-1531579944-adhoc-0:0	6	34	28	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15
A6DFR-1531579944-adhoc-0:0	7	34	28	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15
A6DFR-1531579944-adhoc-0:0	8	34	29	KIAH	BIKF	2018-07-14	14:52	2018-07-14	14:52	2018-07-14	15

Data reduced from 7.83 million obs. with 20 variables (1.5gb) to 1.17 million obs. with 12 variables (138mb)

IDLEWYLD

Training/Testing Split

80% of flights for each **Origin-Destination Pair** (ODP) were used in the training set.

The data set contained 395 Origin-Destination pairs with ≤ 2 flights.

All flights from such pairs were moved to the training set as there wouldn't be enough historical data for both the training and testing processes.

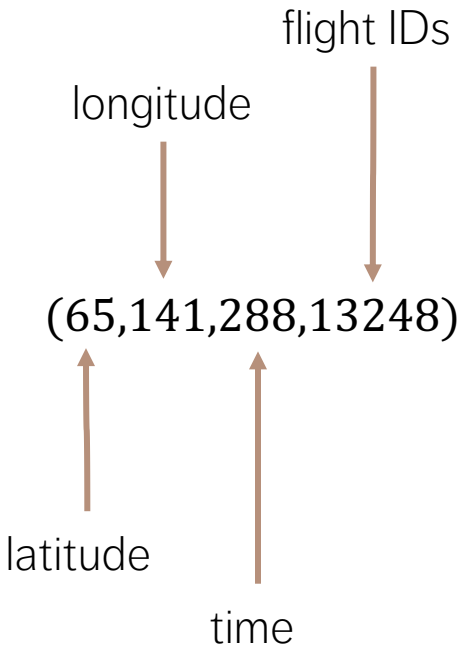
Delay Predictions

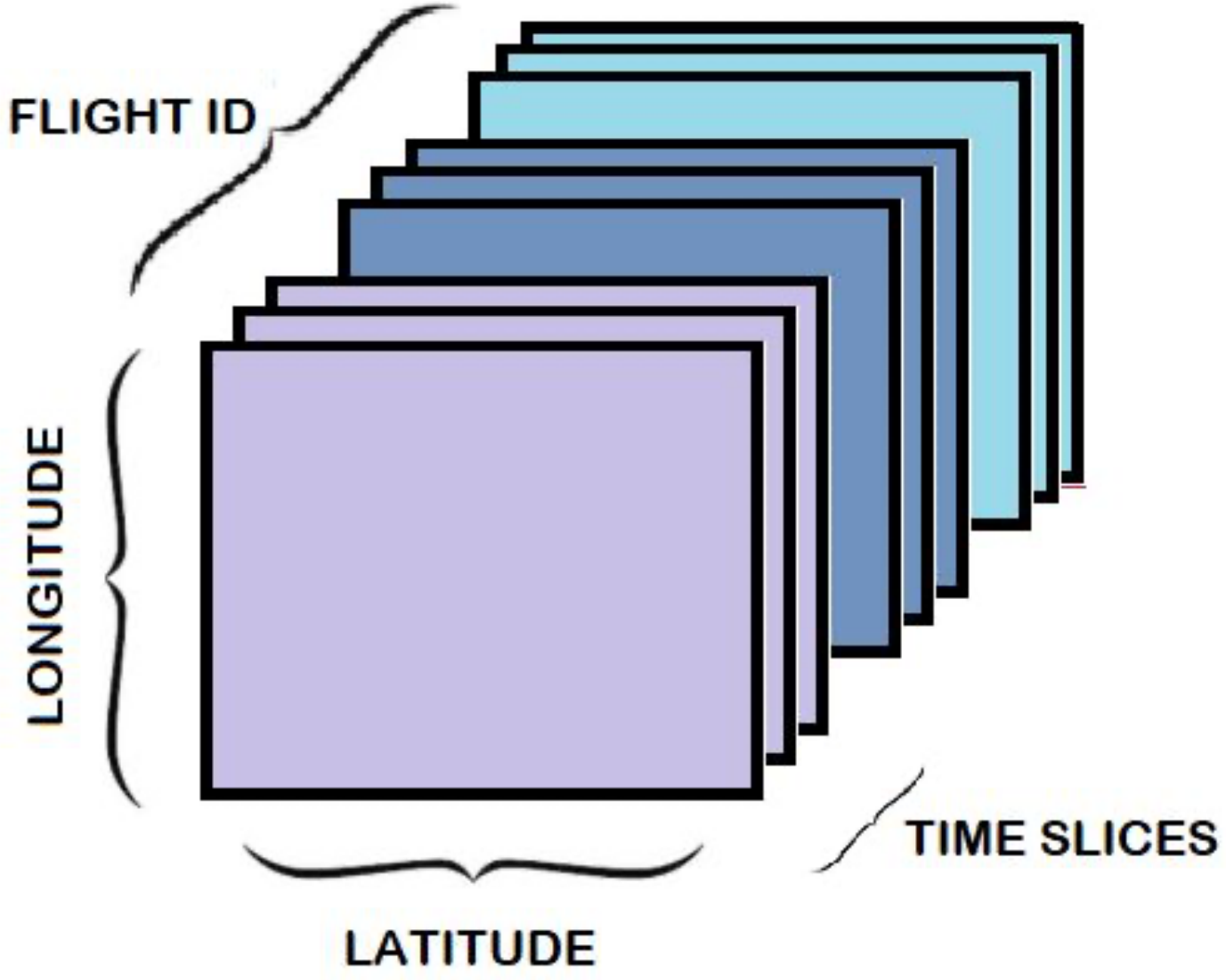
Using “historical” data, we predict the delay for any scheduled flight on a given day.

The schedule specifies both the origin-destination pair (ODP) and the scheduled time of departure

At this stage, delays for each scheduled flight is predicted using a naïve approach (using the distribution of delays for all the flights departing in the same hour for the corresponding ODP).

Tensor

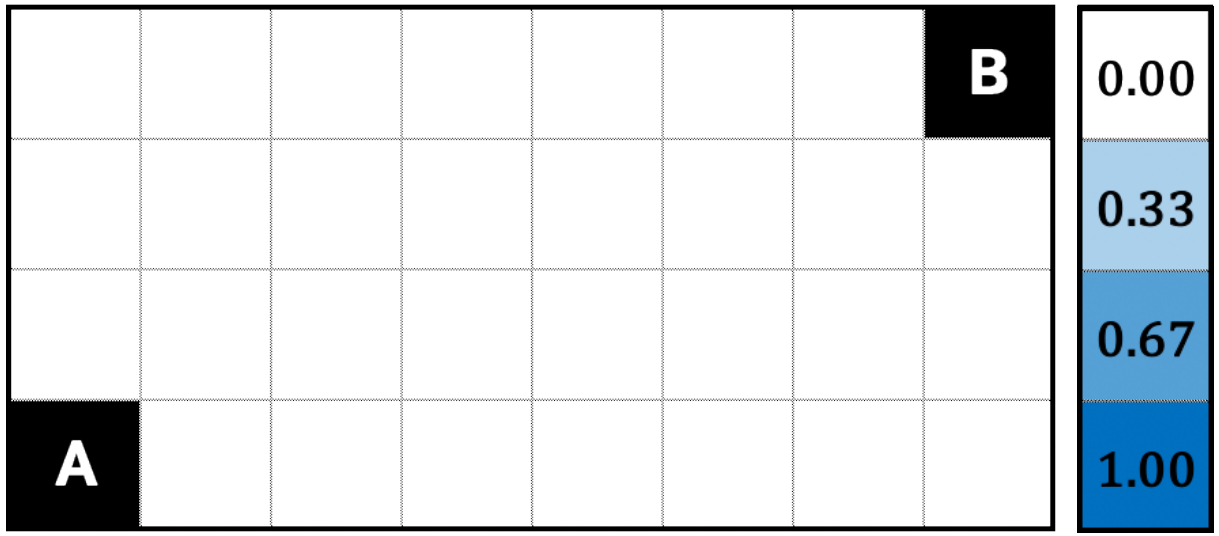




There were 1833 unique ODPs in the dataset.

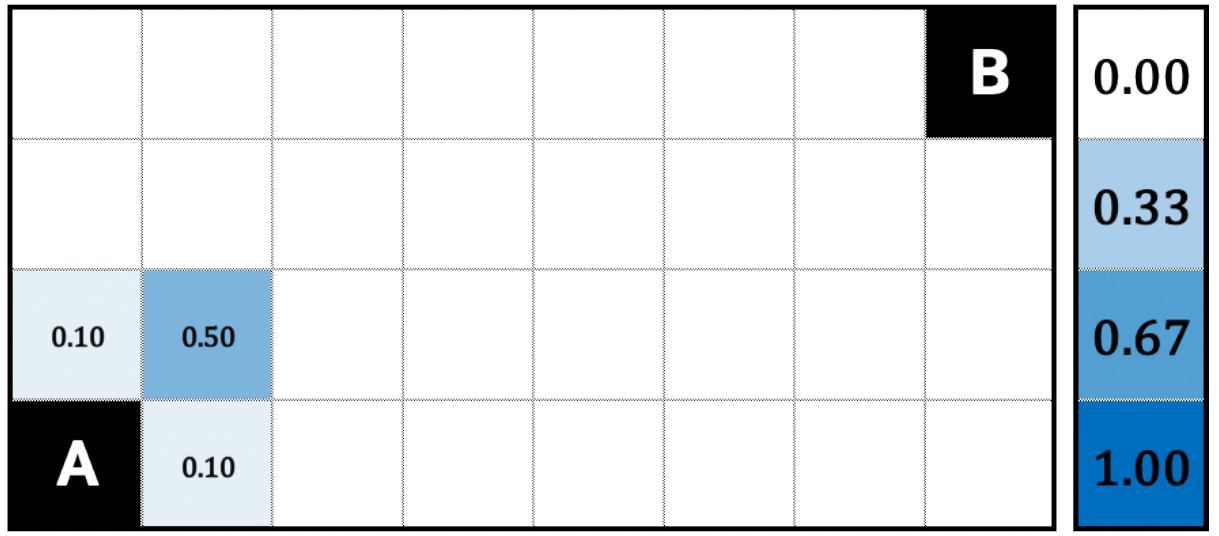
IDLEWYLD

Time Index N



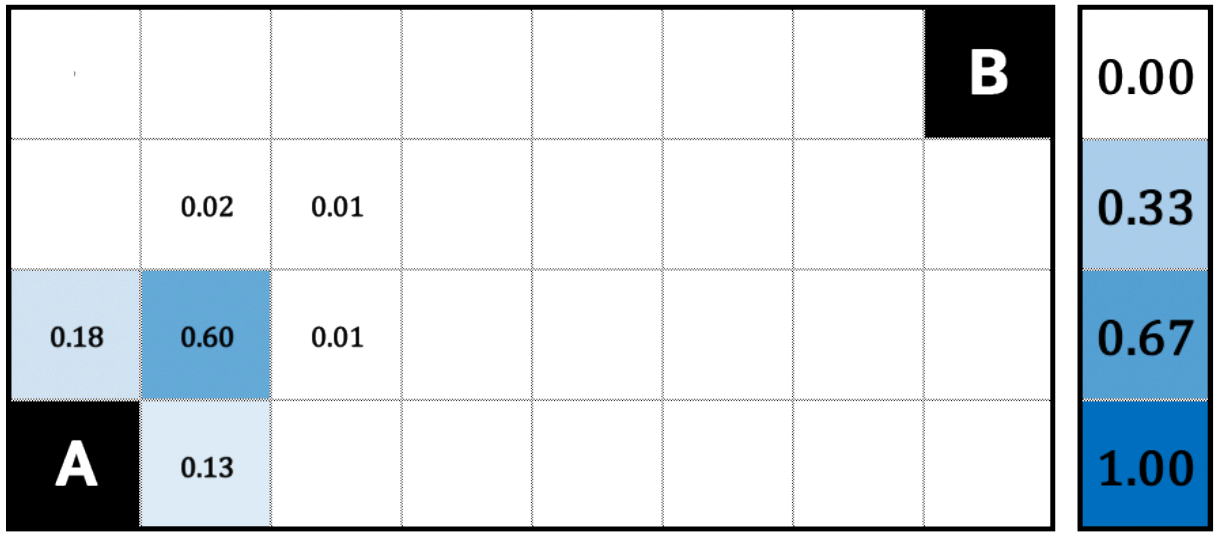
0%

Time Index N+1



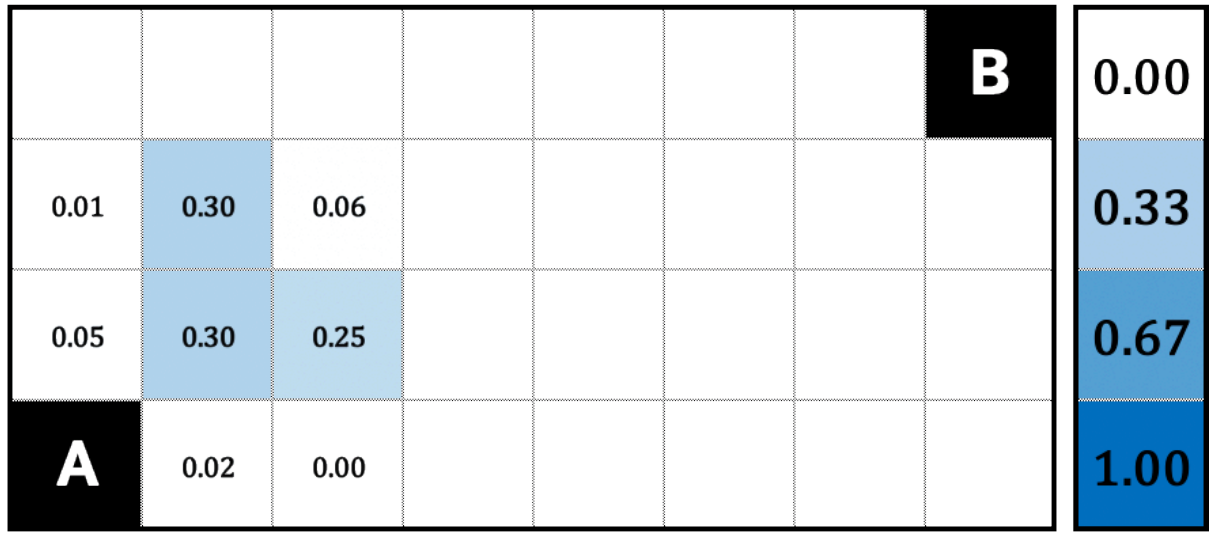
70%

Time Index N+2



95%

Time Index N+3

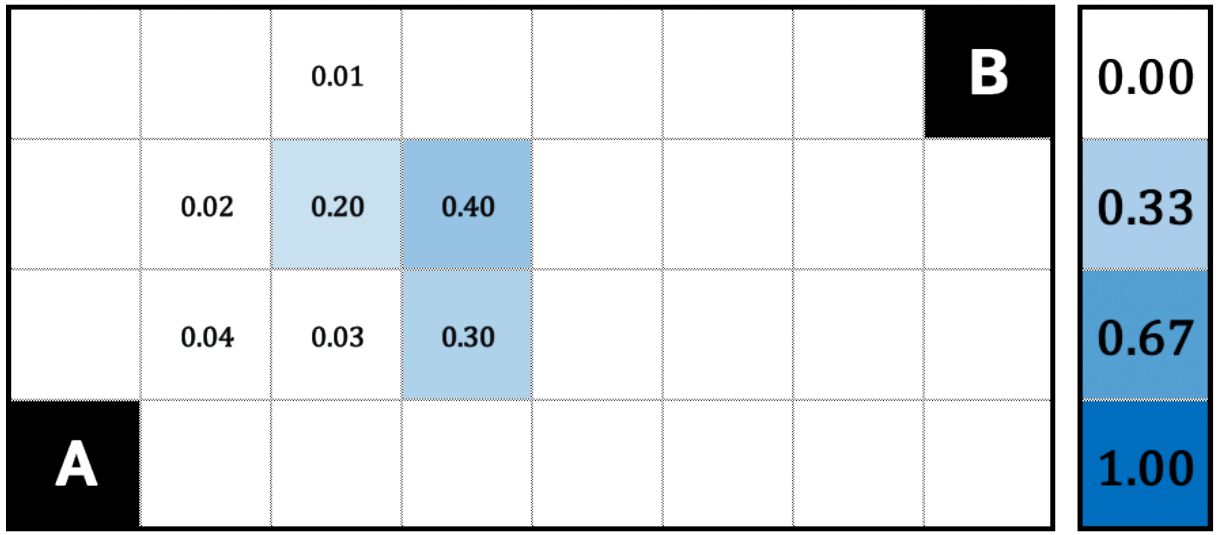


99%

IDLEWYLD

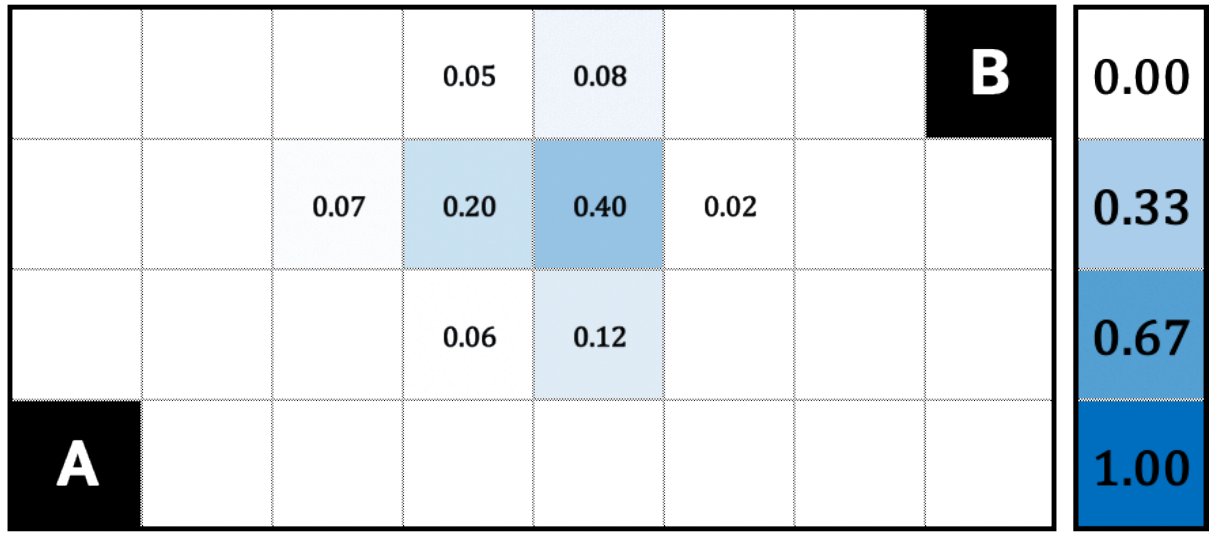

Time Index N
passage of time
Time Index N+1
B
A
0.00
0.33
0.67
1.00
0%
0.10
0.50
0.10
% of flights in that cell at that time
% of flights in the air
70%
Time Index N+2
0.02
0.01
0.18
0.60
0.01
0.13
95%
Time Index N+3
0.01
0.30
0.06
0.05
0.30
0.25
0.02
0.00
99%
IDLEWYLD

Time Index N+4



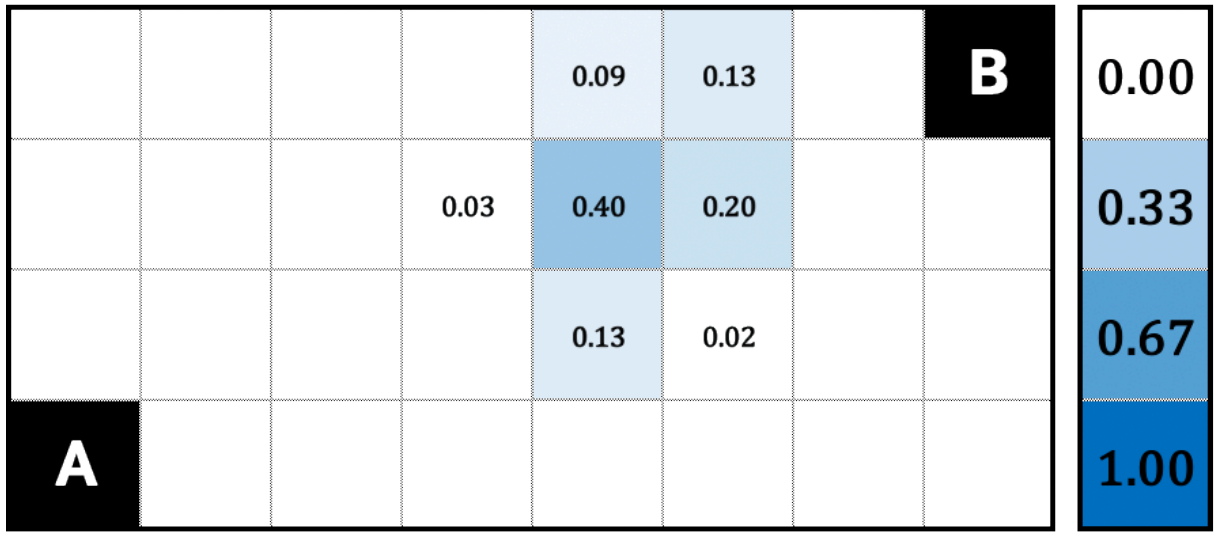
100%

Time Index N+5



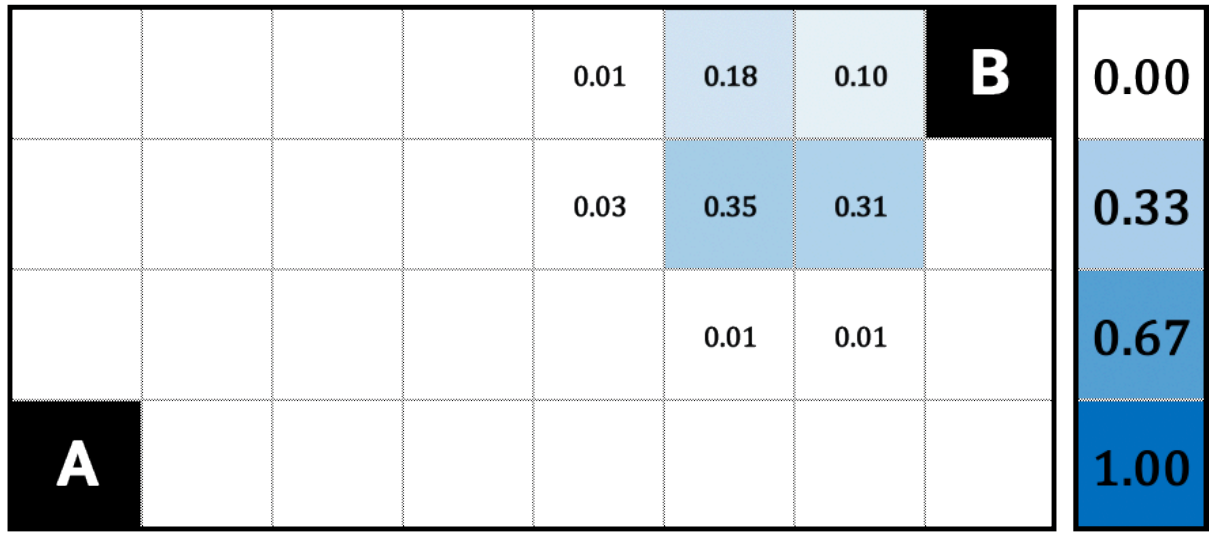
100%

Time Index N+6



100%

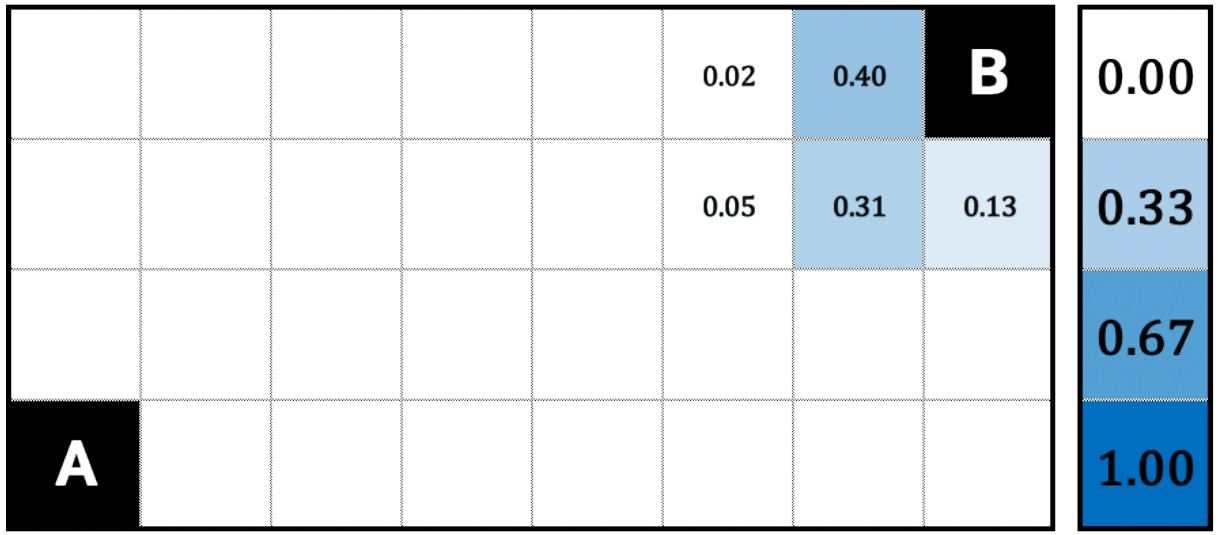
Time Index N+7



100%

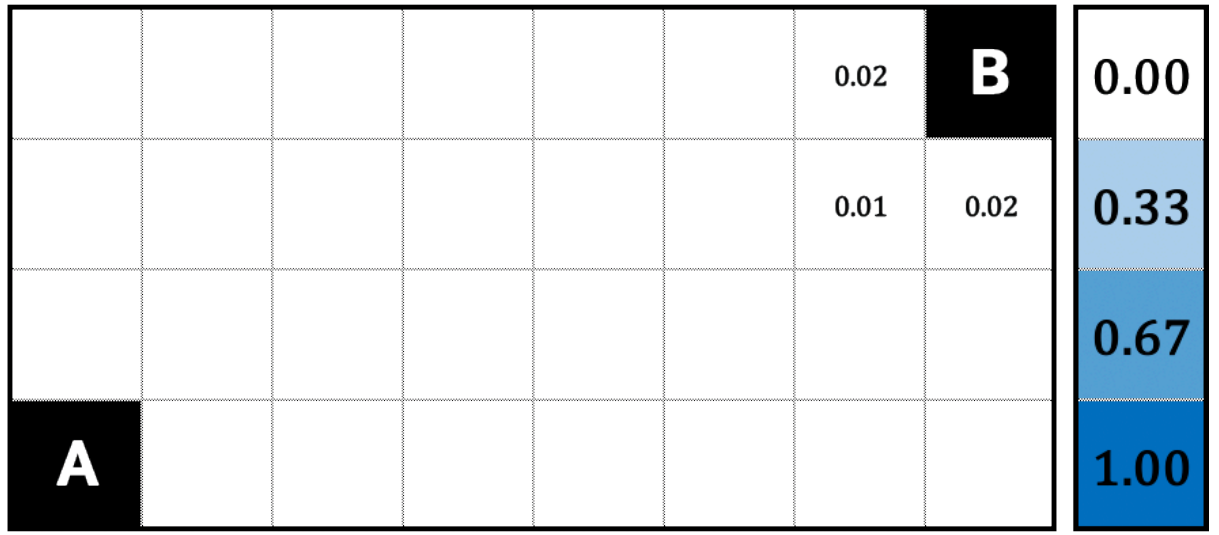
IDLEWYLD

Time Index N+8



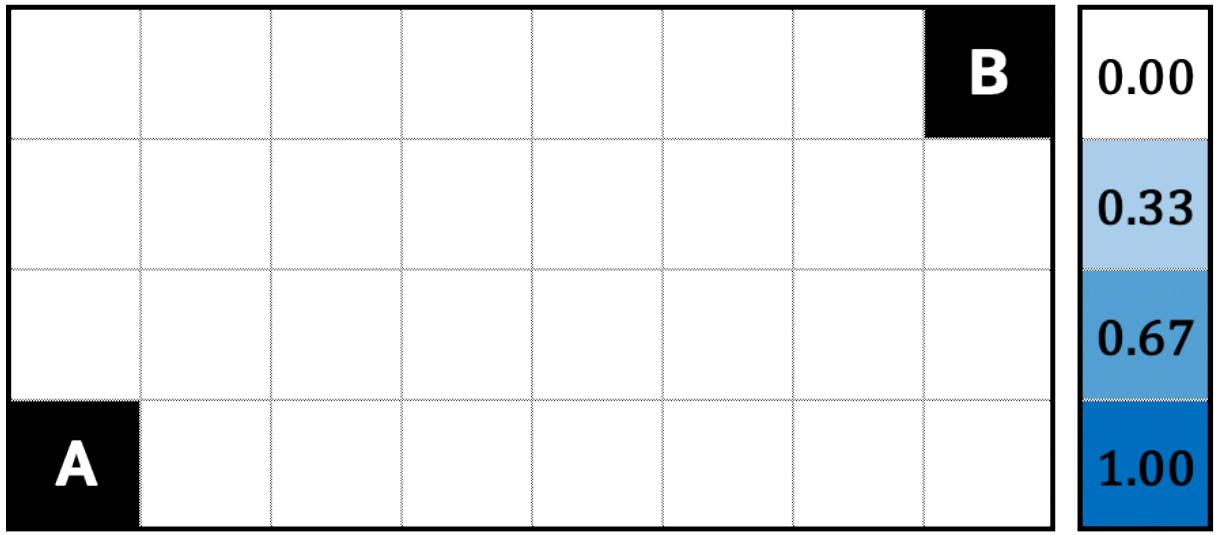
91%

Time Index N+9



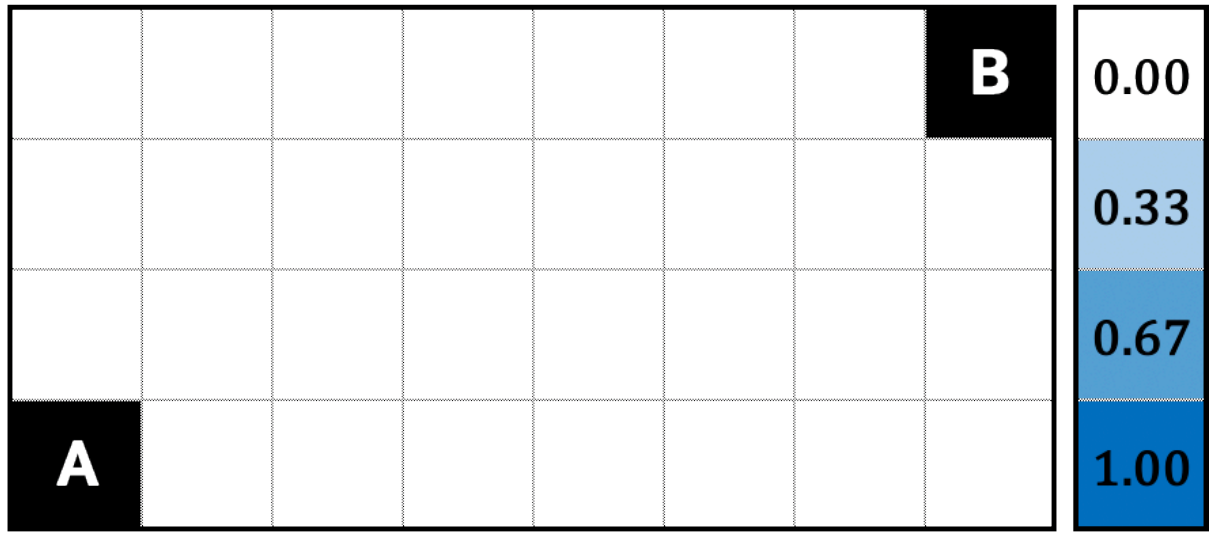
5%

Time Index N+10



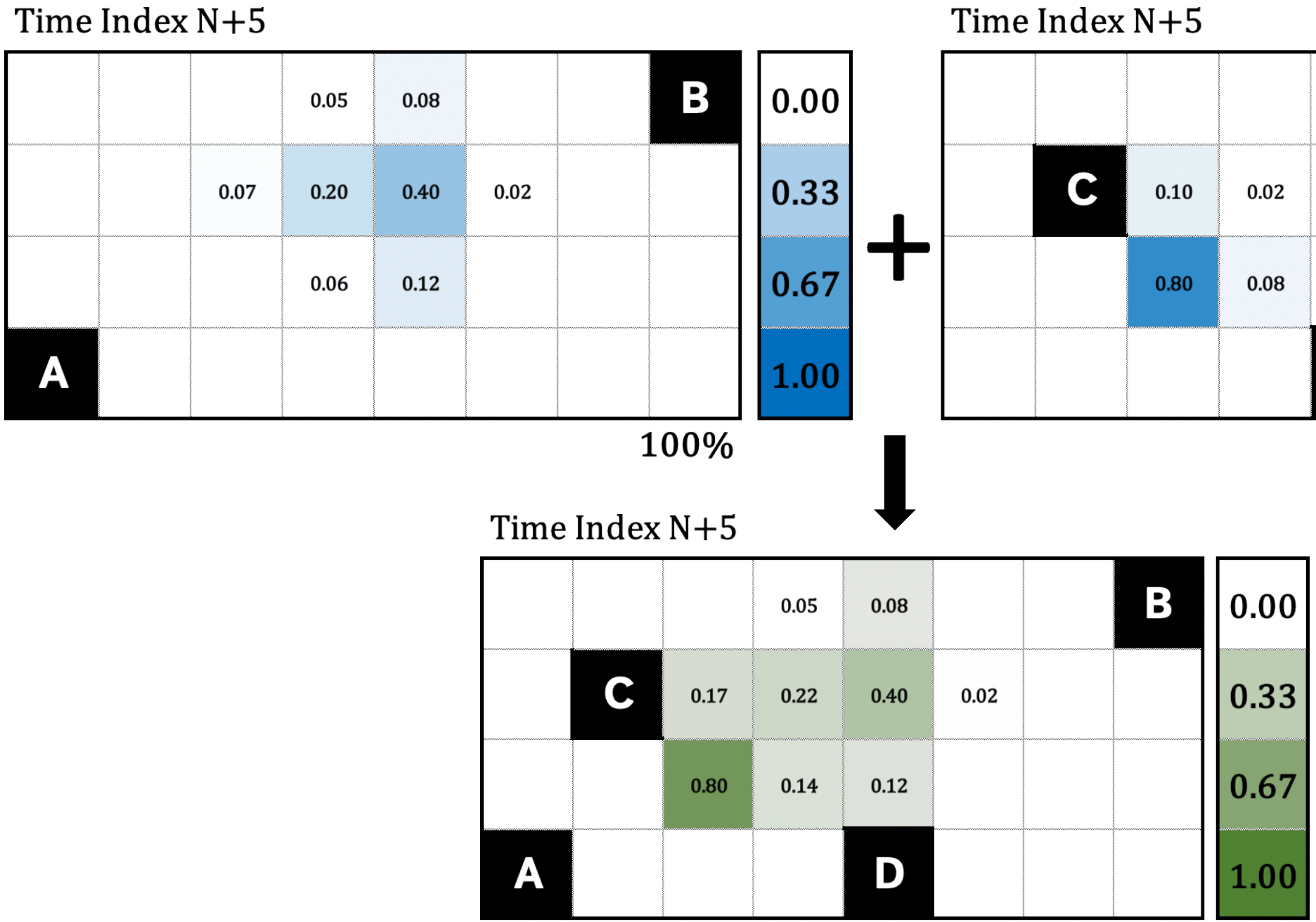
0%

Time Index N+11



0%

IDLEWYLD


Time Index N+5
0.05
0.08
B
0.07
0.20
0.40
0.02
0.06
0.12
A
0.00
0.33
0.67
1.00
100%
Time Index N+5
C
0.10
0.02
0.80
0.08
D
0.00
0.33
0.67
1.00
100%
Time Index N+5
0.05
0.08
B
C
0.17
0.22
0.40
0.02
0.80
0.14
0.12
A
D
0.00
0.33
0.67
1.00
2

IDLEWYLD

Heat Map – Flights Distribution

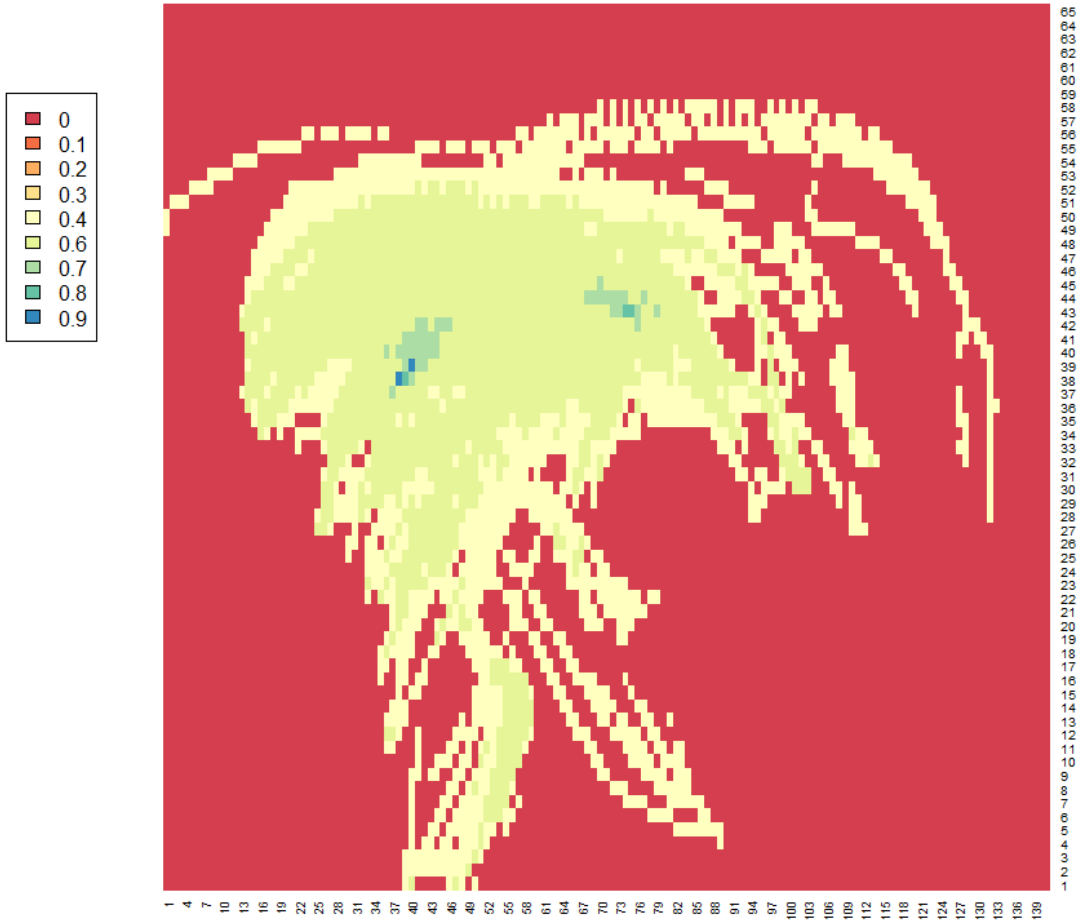
Given a daily schedule, the algorithm generates a heat map displaying where planes are likely to be found during each time period

With such heat maps, we can also identify the regions of highest flight densities, and the estimated number of planes in each region for the day under consideration.

Validation

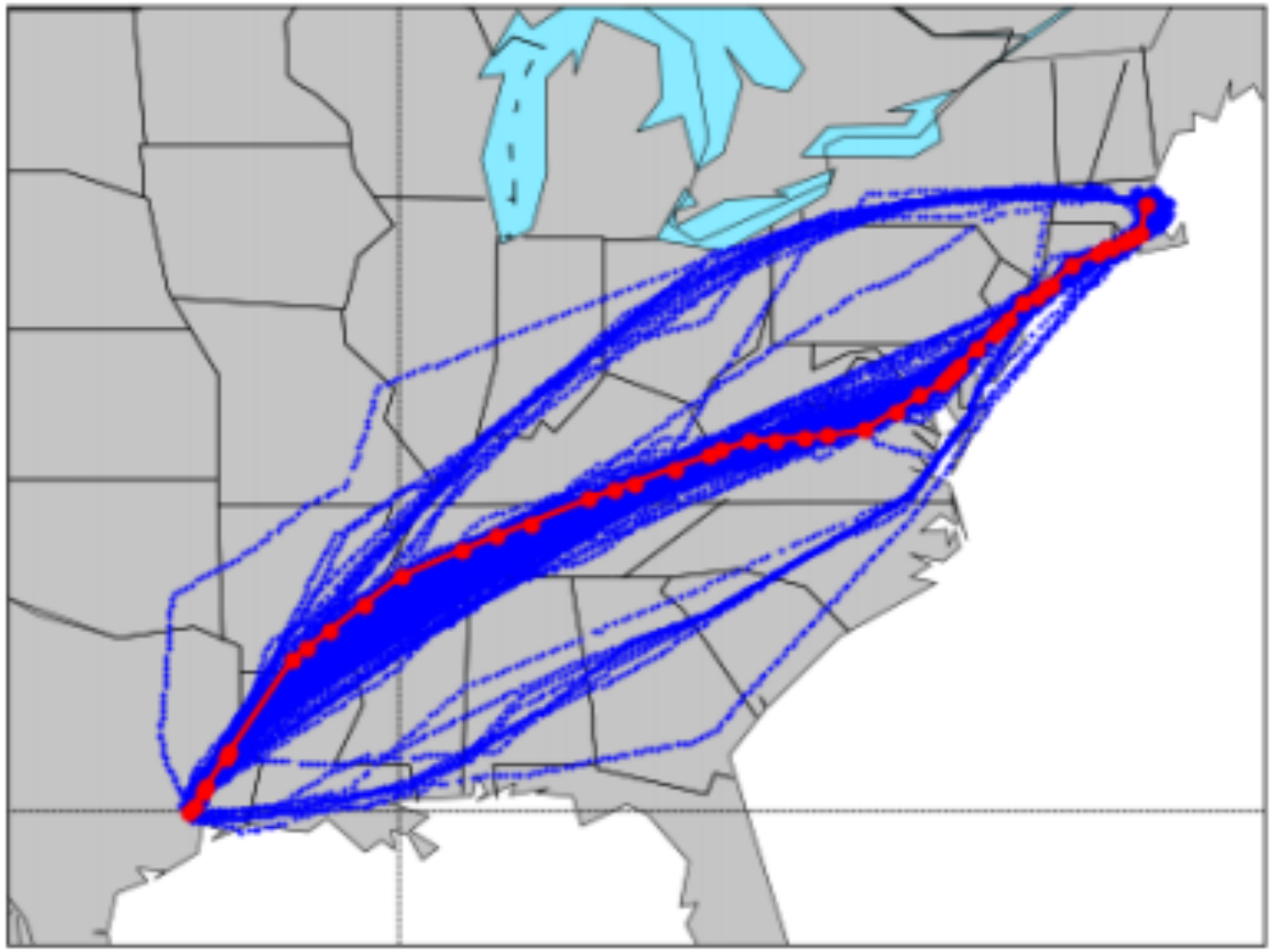
Average Delay Prediction Error ≈ 48min

MAE for Flight Distribution ≈ 0.83



[Liu, Hansen]

Flight Paths vs Heat Maps

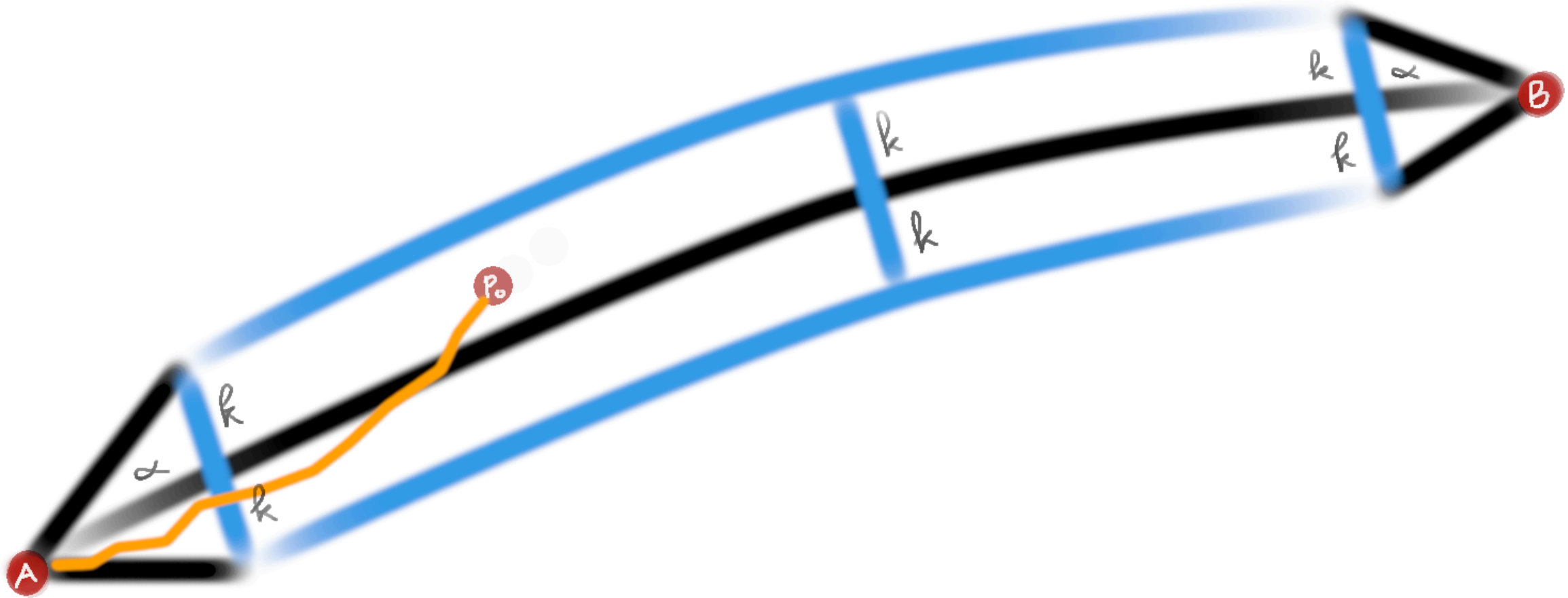


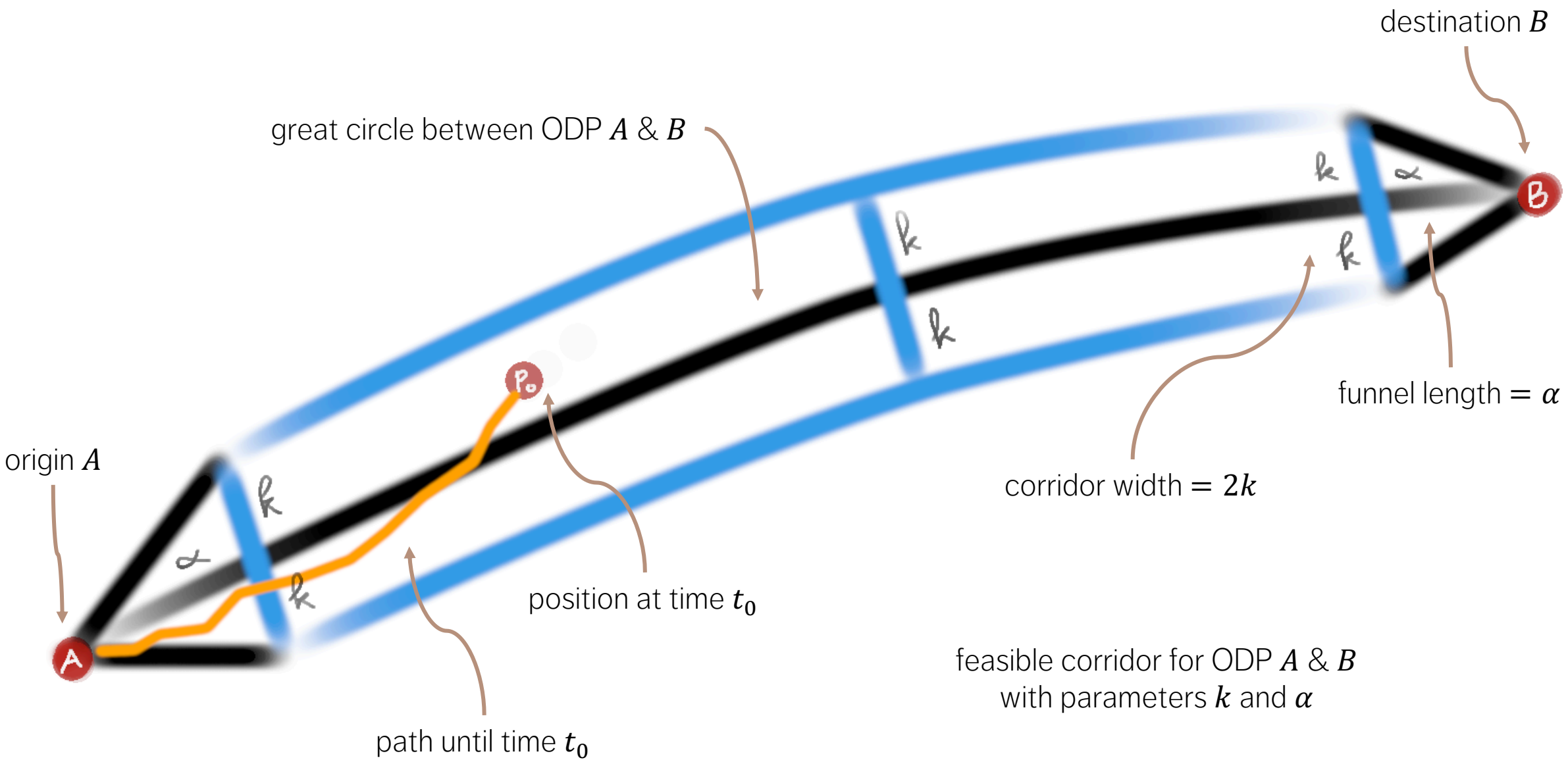
following a specific path

vs.

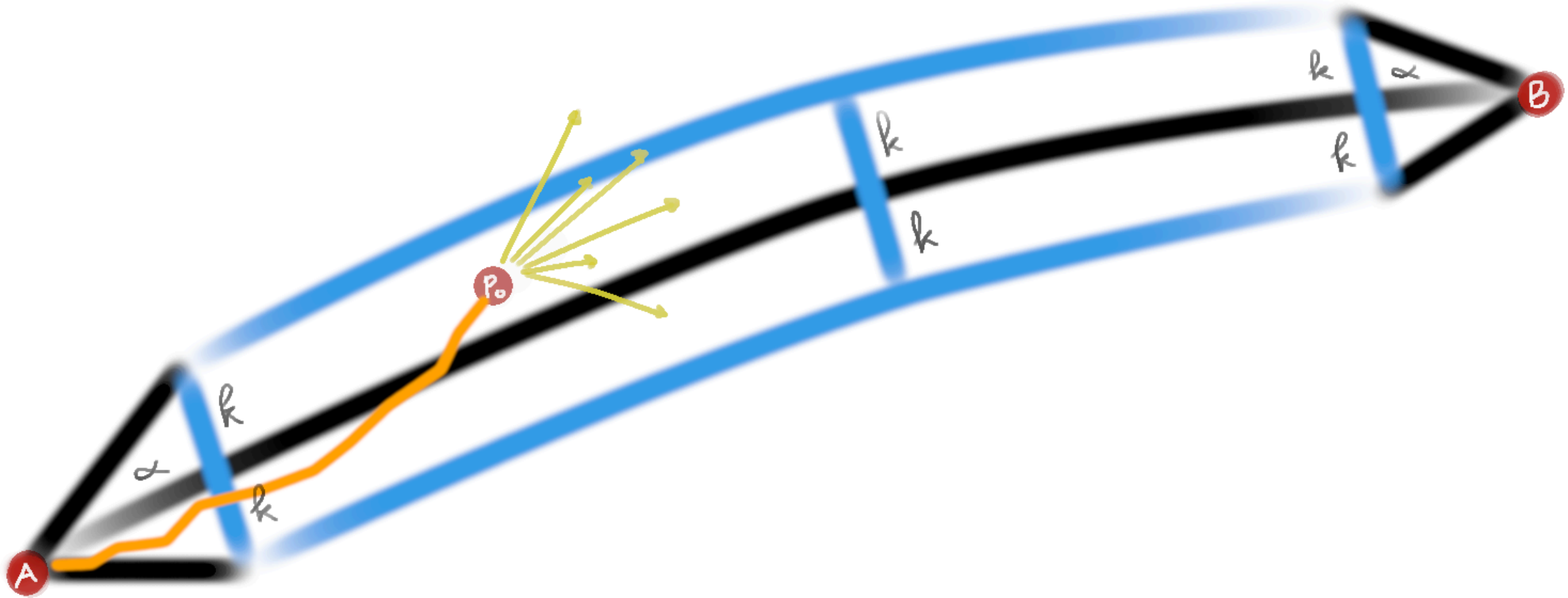
following all potential paths simultaneously

IDLEWYLD



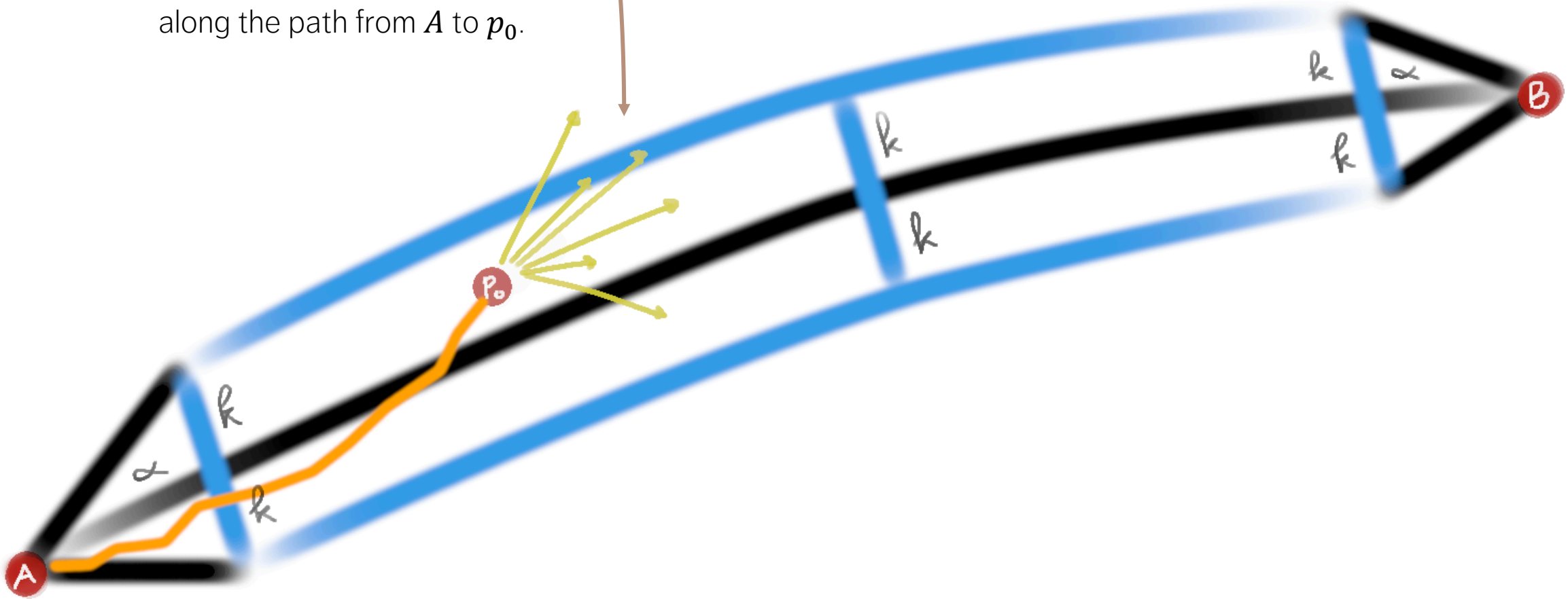

destination B
great circle between ODP A & B
origin A
position at time t_0
path until time t_0
corridor width $= 2k$
funnel length $= \alpha$
feasible corridor for ODP A & B
with parameters k and α
A
B
P_0
k
α

IDLEWYLD

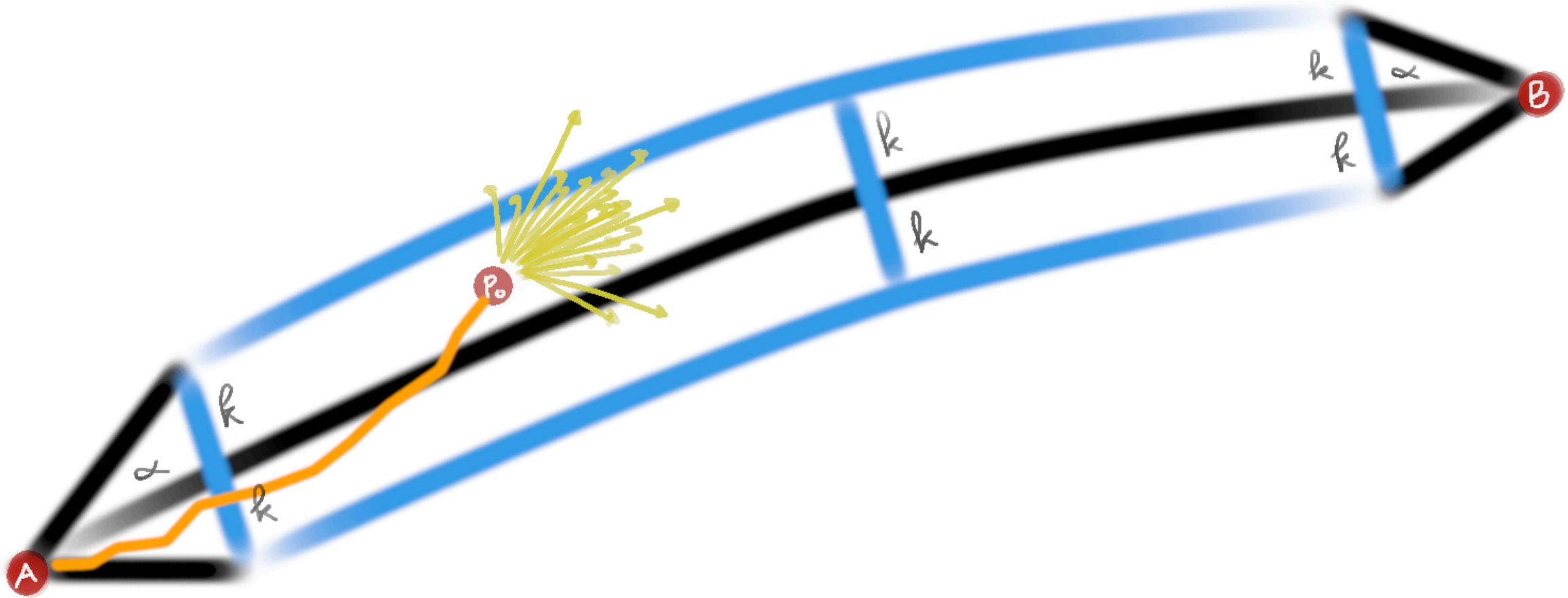


IDLEWYLD

a few samples from the distribution of new positions at time $t_1 = t_0 + \Delta t$, based on position at time t_0, speed s_0, bearing θ_0, and other measurements along the path from A to p_0.

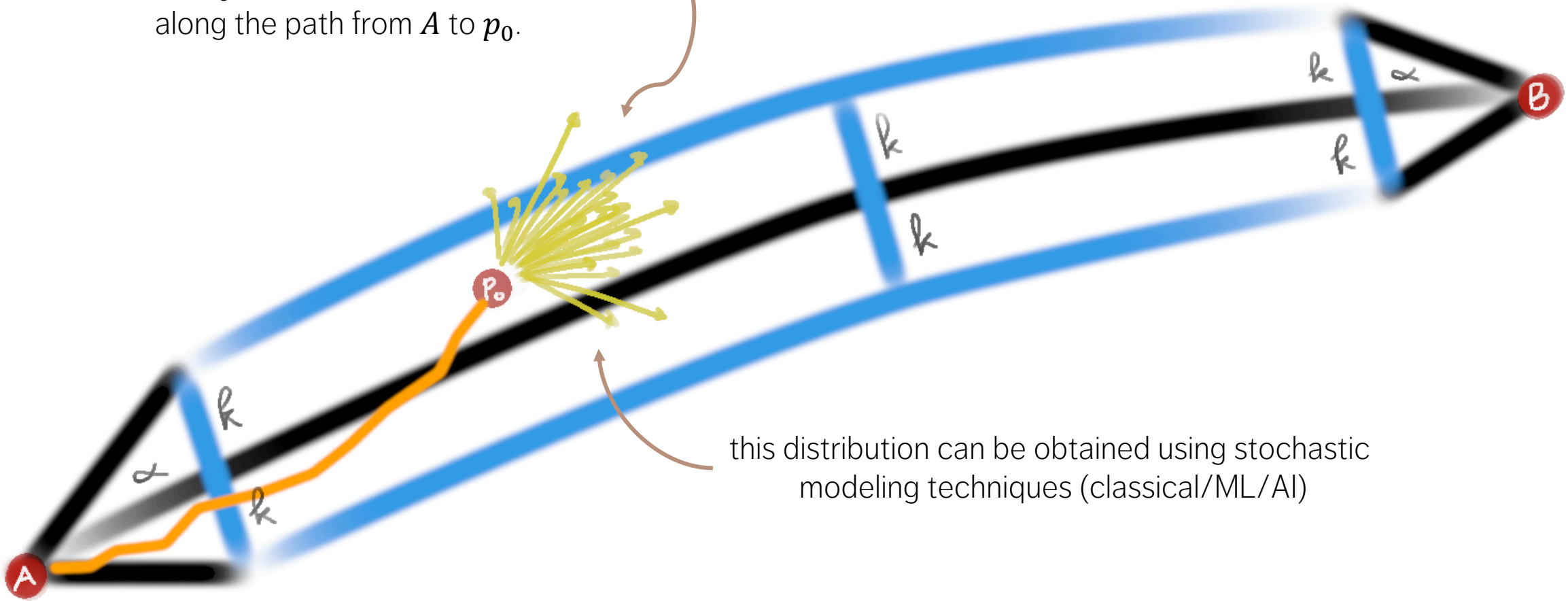


IDLEWYLD



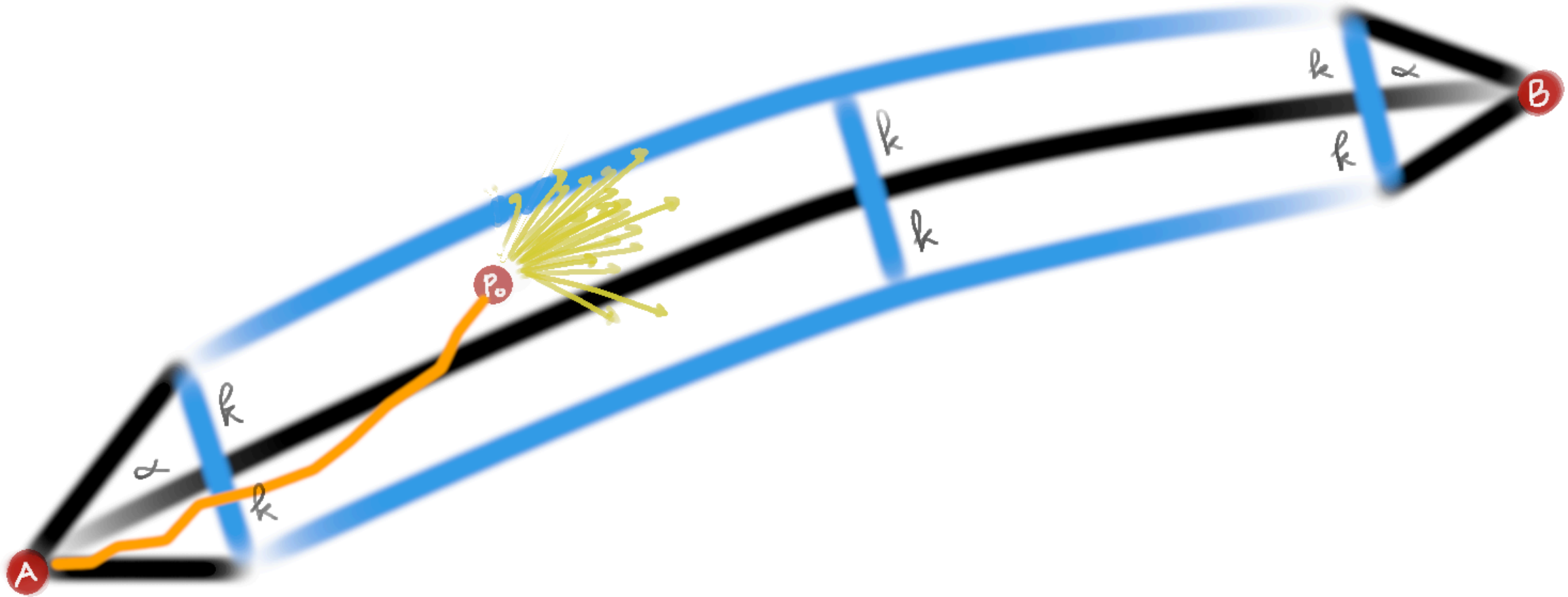
IDLEWYLD

a more complete representation of the distributions of positions at time $t_1 = t_0 + \Delta t$, based on position at time t_0, speed s_0, bearing θ_0, and other measurements along the path from A to p_0.



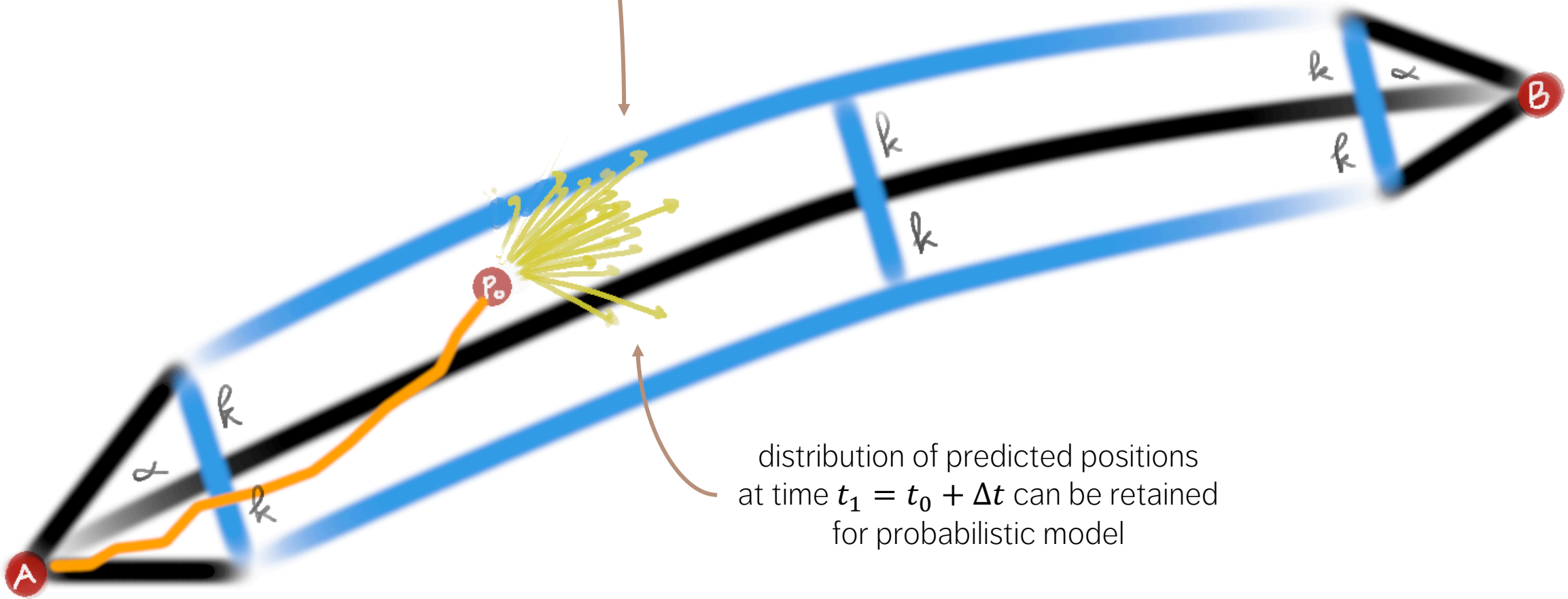
this distribution can be obtained using stochastic modeling techniques (classical/ML/AI)

IDLEWYLD



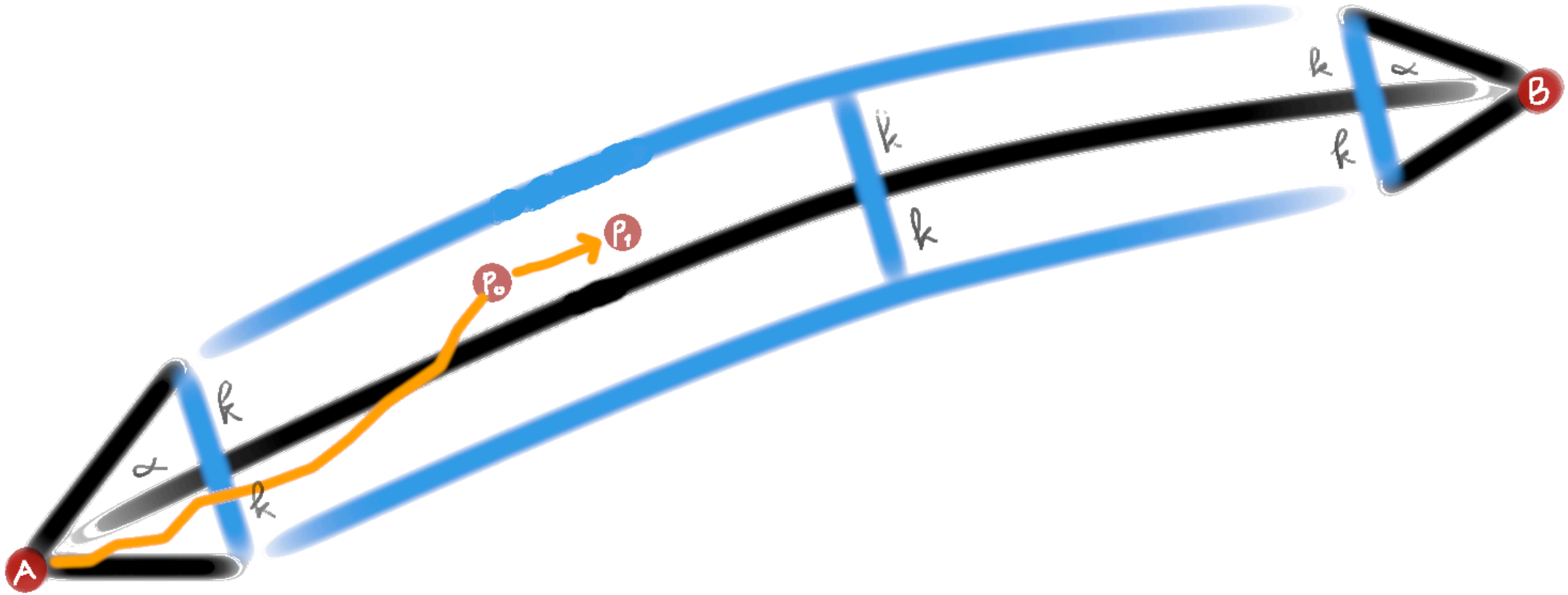
IDLEWYLD

removal of predicted positions
at time $t_1 = t_0 + \Delta t$ falling outside
the feasible corridor (pilot correction)



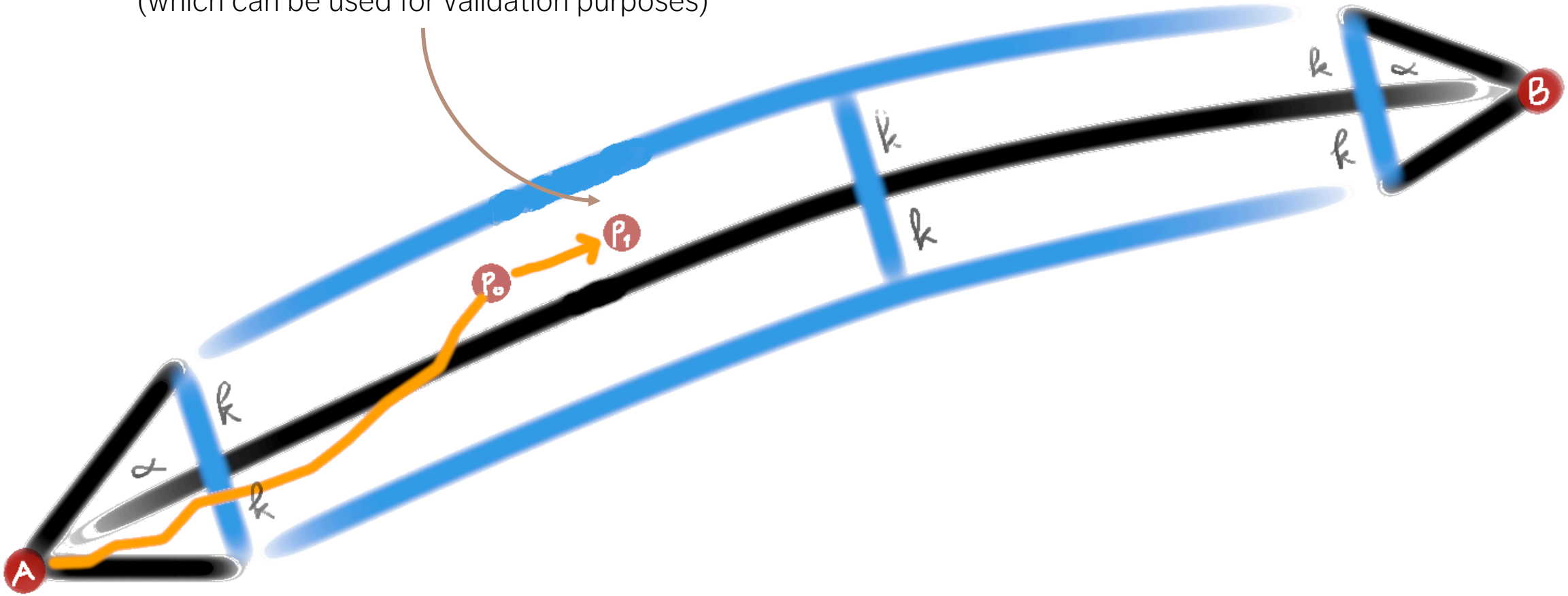
distribution of predicted positions
at time $t_1 = t_0 + \Delta t$ can be retained
for probabilistic model

IDLEWYLD



IDLEWYLD

punctual prediction of position at time $t_1 = t_0 + \Delta t$
(which can be used for validation purposes)



IDLEWYLD

Next Steps – Location Clust.

Stratify each flight by day of the week/month

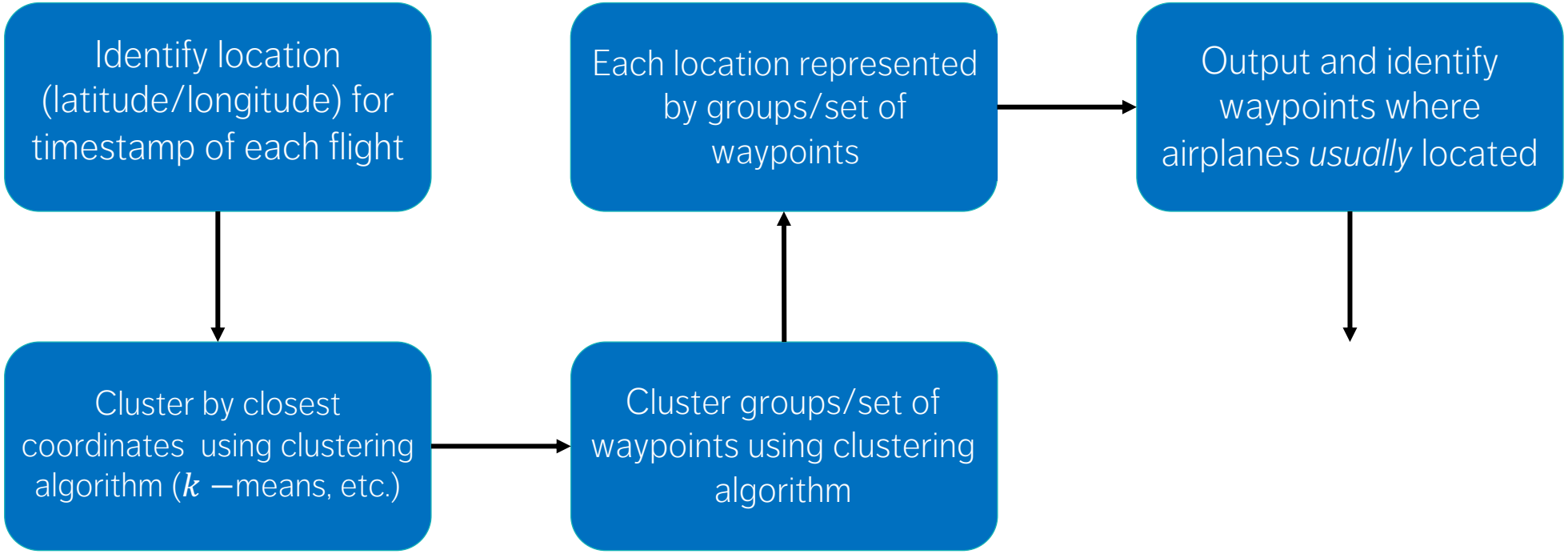
Create 3D-tensor of actual aircraft locations for each day in the historical dataset (similar to heat map)

Model:

- input: day of the week/month
- compare all images or 2D tensors of the actual aircraft location for each time interval for the given day and find “best fit”
- stitch the forecasts for each time interval back into a 3D tensor

IDLEWYLD

Next Steps – Waypoint Clust.



IDLEWYLD

TRANSPORTATION CLUSTERS

3.200 airports
60.000 routes

GEOGRAPHICAL LAYOUT

Color = Longitude
Size = Number of routes

EUROPE
Alaska
Canada
Russia
NORTH AMERICA
ASIA
AFRICA
India
MIDDLE EAST
LATIN AMERICA
Pacific islands
OCEANIA

Data: openflights.org

Next Steps – Misc.

Other models under consideration:

> Predicting flight routes with a Deep Neural Network in the operational Air Traffic Flow and Capacity Management system:

Predicting when flights are cancelled (requires historical data).

Incorporating spherical geometry and flight altitude (mathematically feasible, but could prove conceptually problematic).

Improving validation procedure (will require more training data).