



プロセスマイニング技術を応用した空港面交通流のメソスコピックモデル化手法

上原 健嗣 平石 邦彦

北陸先端科学技術大学院大学 情報科学系 平石研究室

CARATSオープンデータ活用促進フォーラム
2021年12月13日

1. 概要

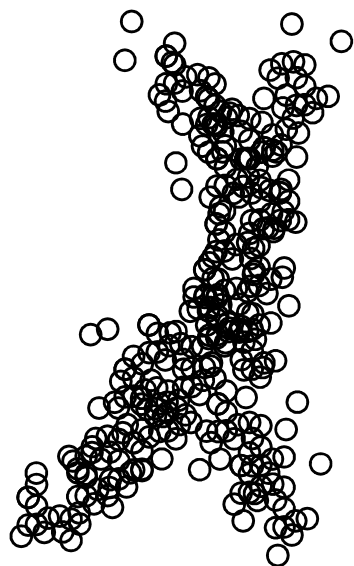
巨大な交通流データをモデル化するにあたり、必要な情報粒度は確保しつつも、データサイズを縮退させた**メゾスコピックモデル**を考える。本稿では、**プロセスマイニング技術**を応用し、メゾスコピックモデルを効率的に作成する手法を紹介します。

主な内容：

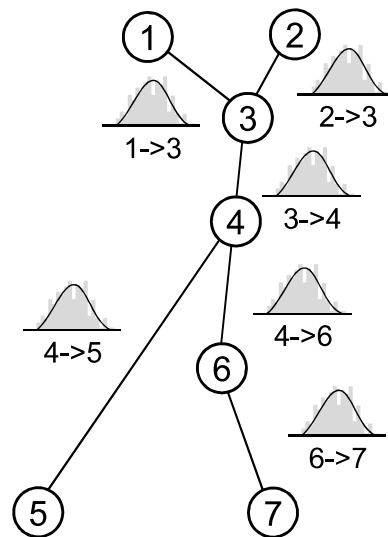
- メゾスコピックモデルとは
- プロセスマイニングとは
- CARATS空港面交通流データを用いたモデリングデモ

2. メゾスコピックモデルとは

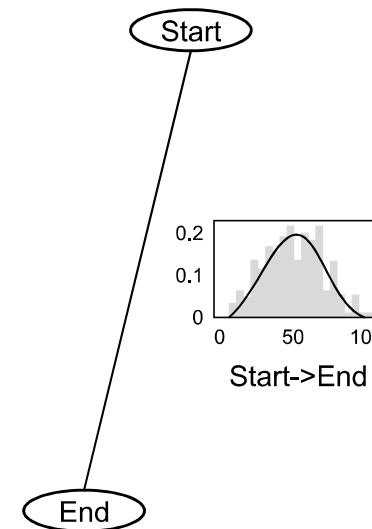
解析者が何を知りたいかによって、必要なデータ粒度は変わります。
ミクロほど細かすぎず、マクロほど粗すぎず、その**中間の粒度**をメゾスコピックと言います。



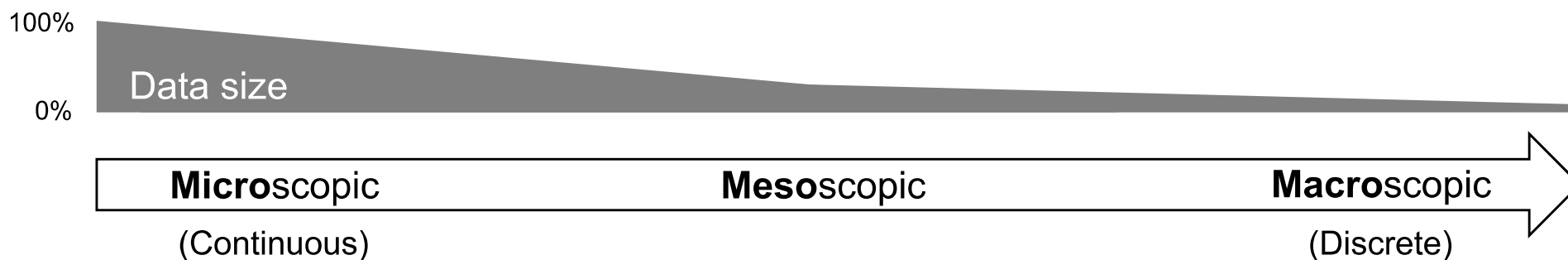
Empirical raw data



Abstracted network

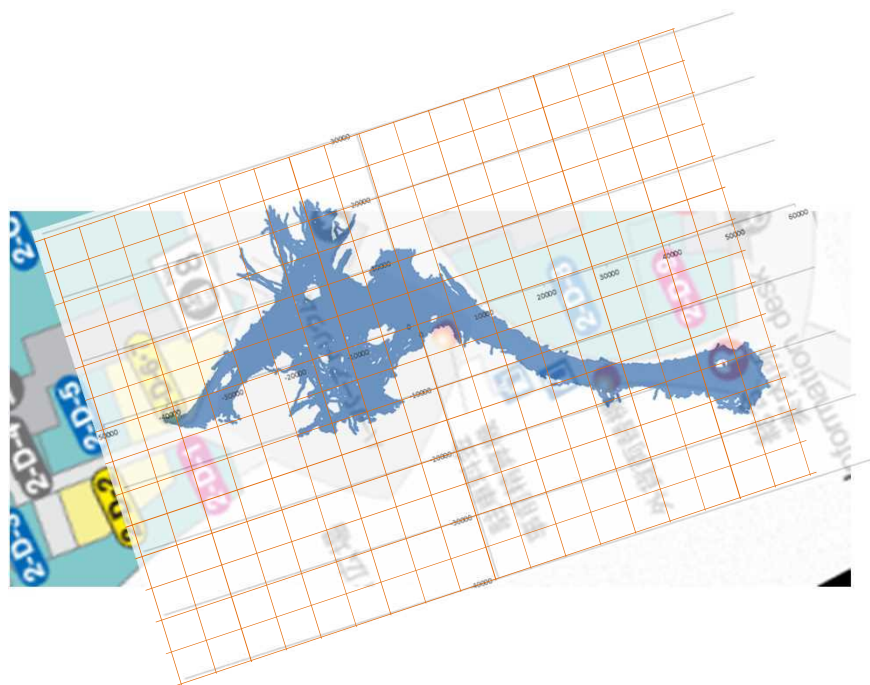


Simplified start-to-end graph

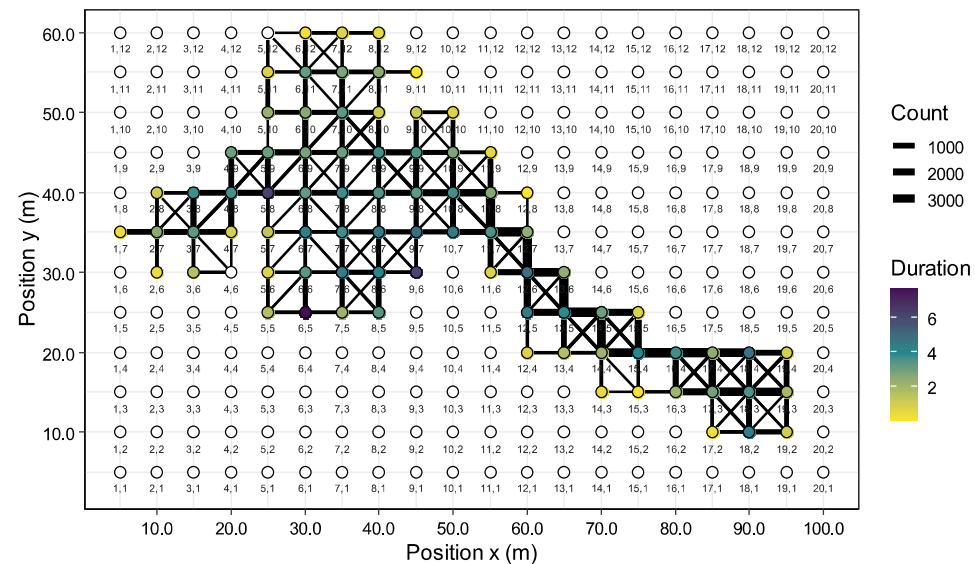


3. メゾスコピックモデルの例

以下は、歩行者の挙動データをメゾスコピックのレベルまで抽象化した例です。



抽象化



使用した歩行者データセットには、1日分で約18000人分の動態情報が格納されており、サイズは1GBを超える。これを格子で区切り、各格子（ノード）間における移動情報に置き換える。例として5.0m格子でノードを形成した場合、データ量はわずか8MBとなる。しかしながらこのレベルでも、**人流を把握するだけであれば十分**。

Datasets from the ATC shopping center: http://www.irc.atr.jp/crest2010_HRI/ATC_dataset/

D. Brscic, T. Kanda, T. Ikeda, T. Miyashita, "Person position and body direction tracking in large public spaces using 3D range sensors", IEEE Transactions on Human-Machine Systems, Vol. 43, No. 6, pp. 522-534, 2013

4. メゾスコピックモデルのための抽象化

抽象化のレベル決定は非常に重要。これは「モデルを使って交通流をどのレベルまで知りたいのか？」に全て依存します。

必要な情報を保ったままデータ量を減らす

- ある程度減らす→その「ある程度」とはどのくらいなのか
- 減らしたからと言って、必要な情報まで欠落させてはならない
- どうやって減らすのかがポイント

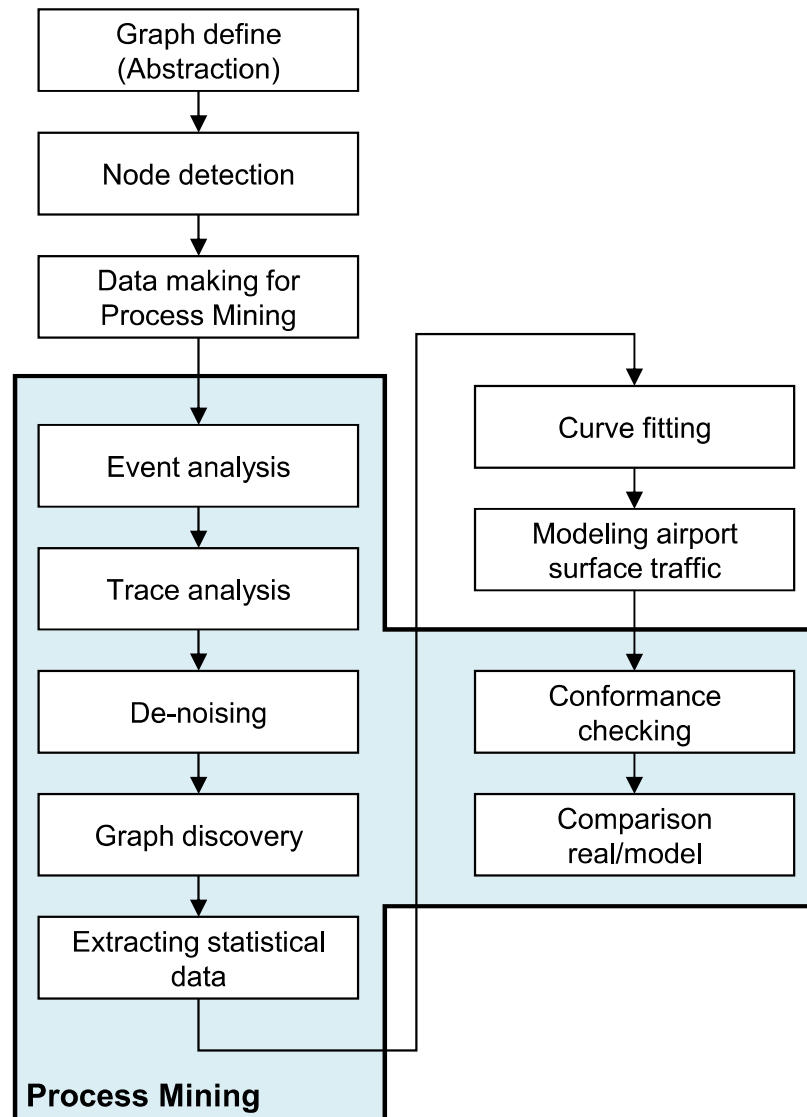
連続的なデータを、移動情報として統計値に置き換える

- 大まかなフローは分かる
- 各移動体（人や車）の速度分布、所要時間分布も得られる
- どこで滞留しているのかも大体分かる

よって「大まかなトラフィックの流れをパッと知りたいが、交通の要所となる地点の混雑状態などもそれなりに知りたい」と言った場合に適しています。

5. 交通流メゾスコピックモデル化の流れ

抽象化を行うために、必要な情報を抽出することがメイン作業となります。

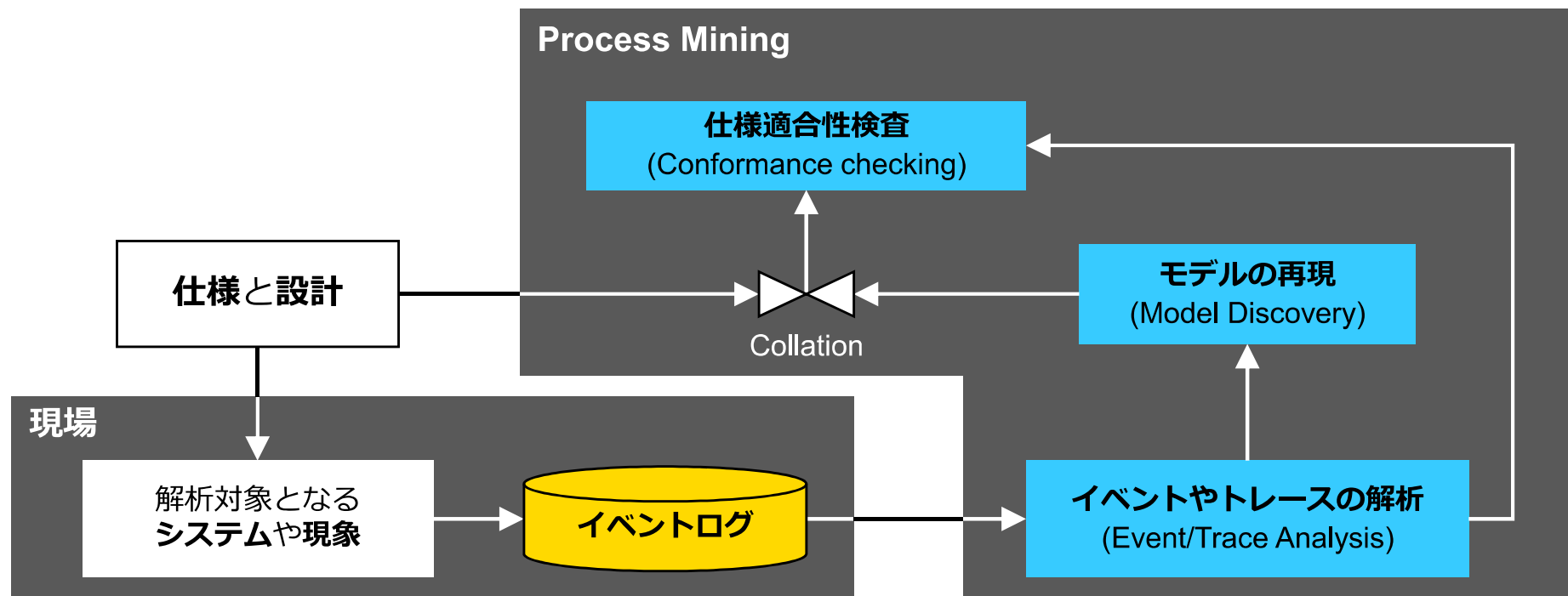


今回は**CARATSオープンデータ**を用いて、サンプルのメゾスコピックモデルを作成します。

CARATSオープンデータのような大容量の交通流の場合、抽象化作業はかなりの労力を要するため、**プロセスマイニング**を用いてこれを効率化します。

6. プロセスマイニングとは

あるシステムの挙動が記録されたイベントログから、有用な情報を抽出・発見する技術。

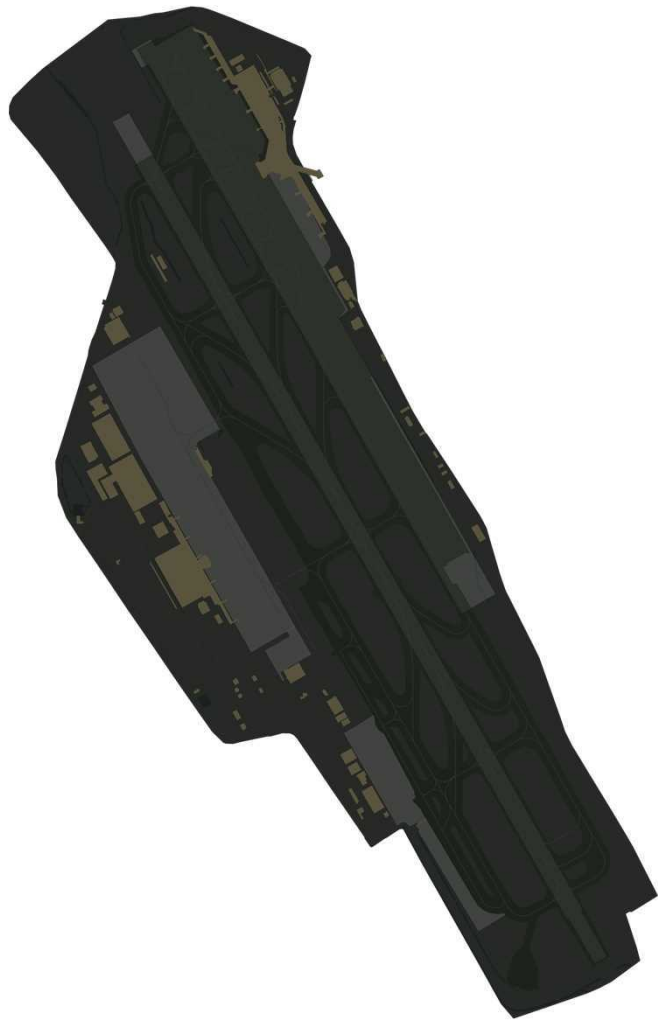


入手可能なデータがイベントログしかないようなシステムでも、厳密な検査ができる。イベントをトリガーとして状態が遷移するようなシステムの解析に適している。

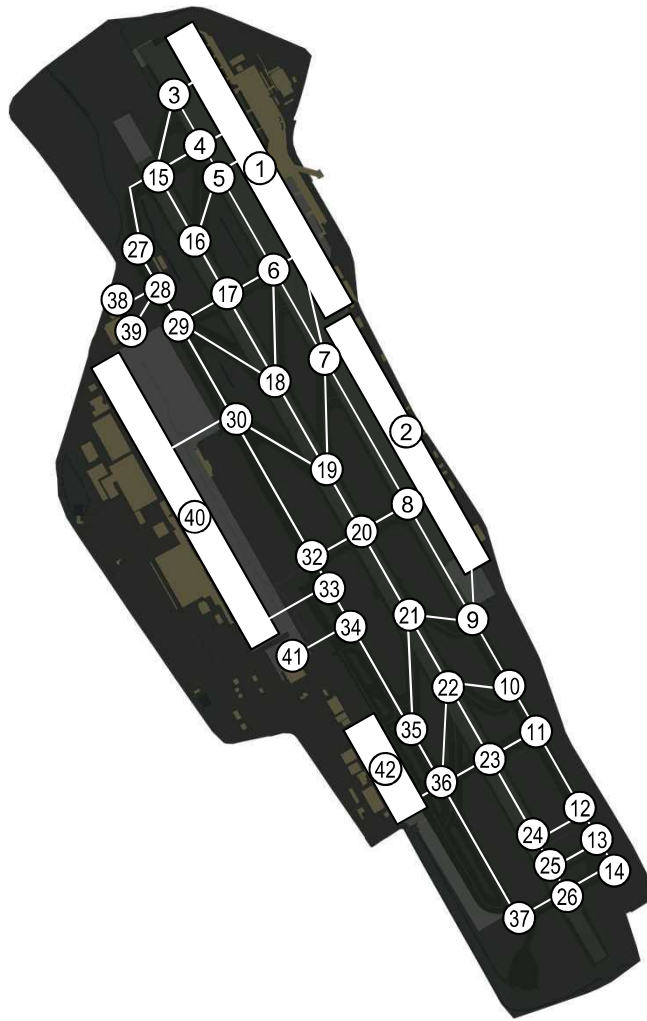
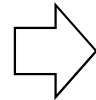
強力な解析ツールではあるが、粒度の高いCARATSデータをそのまま扱うことはできないので、イベントログ化 (= 抽象化) が必要。

7. 空港面交通流の抽象化

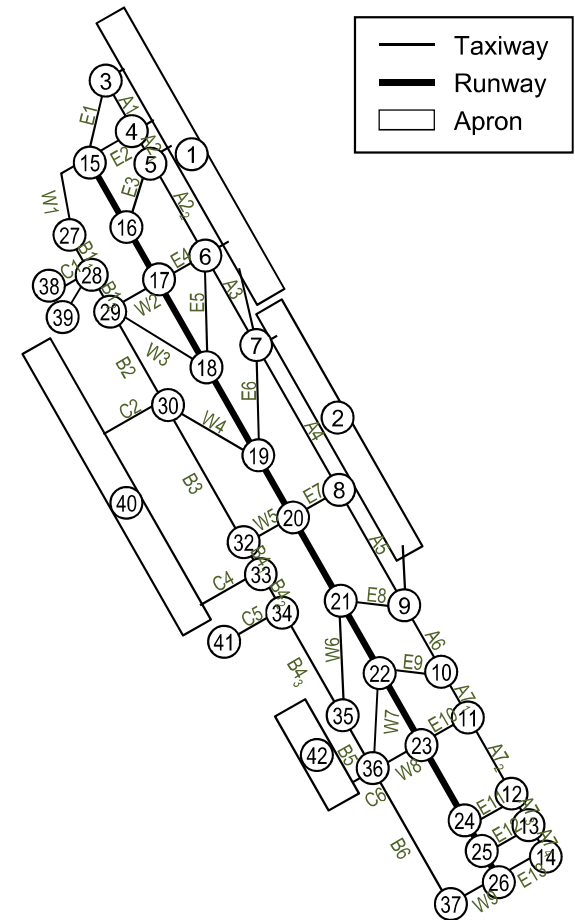
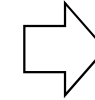
交通の要所が決まっているので**非常に楽** (i.e. 滑走路、誘導路、エプロン、それらの合流点)



Airport Surface (Fukuoka)



Define Edges and Nodes



Graph

8. ノード抽出とイベントログ化

生データに対して全ノード矩形との内外判定を行い、通過情報をイベントとしてロギングします。

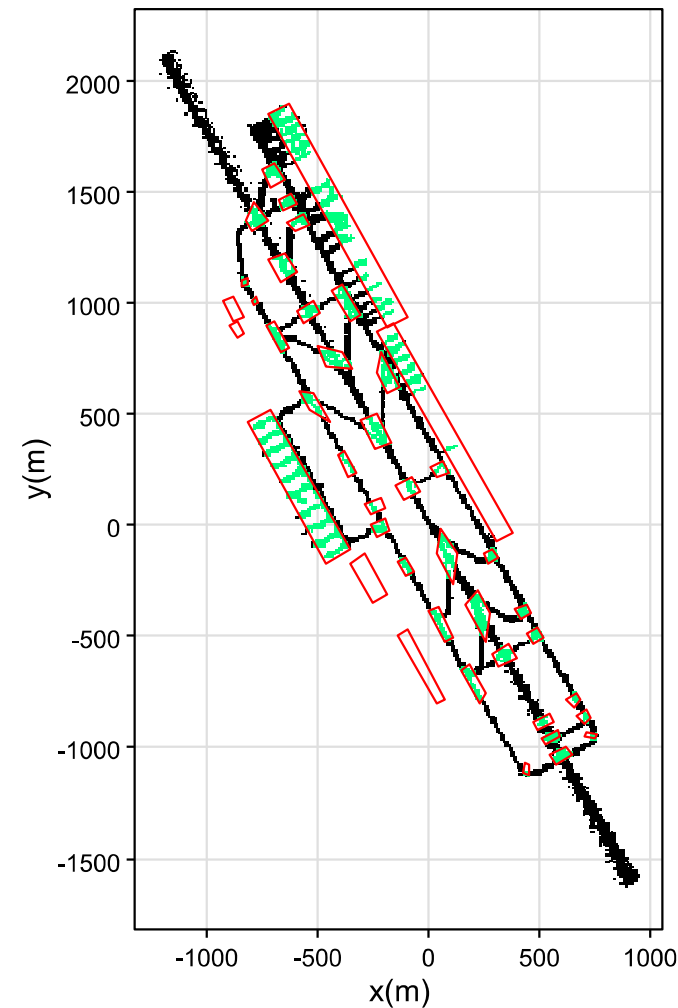
生データ (福岡空港2017)



ノード矩形設定



内外判定後



9. イベントログ化

各航空機が主要ポイントをヒットしたタイミングが、イベントログとして残ります。このレベルまでダウンサイジングしたら、プロセスマイニングが可能になります。

case	org:resource	event	timestamp	transition	lifecycle:transition
1	AP00437	E-APRON	2017-04-17T07:05:13	PT5	complete
1	AP00437	A21	2017-04-17T07:05:40	PT4	complete
1	AP00437	E2	2017-04-17T07:07:04	PT15	complete
1	AP00437	RWY	2017-04-17T07:08:07	PT16	complete
1	AP00437	END	2017-04-17T07:08:57	#	complete
2	AP00438	E-APRON	2017-04-17T07:06:41	PT3	complete
2	AP00438	E1	2017-04-17T07:08:48	PT15	complete
2	AP00438	RWY	2017-04-17T07:09:58	PT16	complete
2	AP00438	END	2017-04-17T07:10:49	#	complete
3	AP00448	E-APRON	2017-04-17T07:11:39	PT5	complete
3	AP00448	A21	2017-04-17T07:12:10	PT4	complete
3	AP00448	E2	2017-04-17T07:12:44	PT15	complete
3	AP00448	RWY	2017-04-17T07:14:27	PT16	complete
3	AP00448	END	2017-04-17T07:15:15	#	complete
...					

イベントログ中のcaseは、航空機毎の通過経路を表し、プロセスマイニングにおいては**トレース**として扱われます。

10. プロセスマイニングツール

プロセスマイニングのツールはオープンソース、商用ソフトの両方があります。ここでは、蘭アイントホーフェン工科大学が提供している**ProM**を使用します。



ProMの特徴

- Java上で動作するオープンソースソフトウェア
- 無償で利用可能
- 頻繁なアップデート
- 豊富なプラグインツール

ProM (Process Mining Workbench)

<https://www.promtools.org/>

11. イベント分析

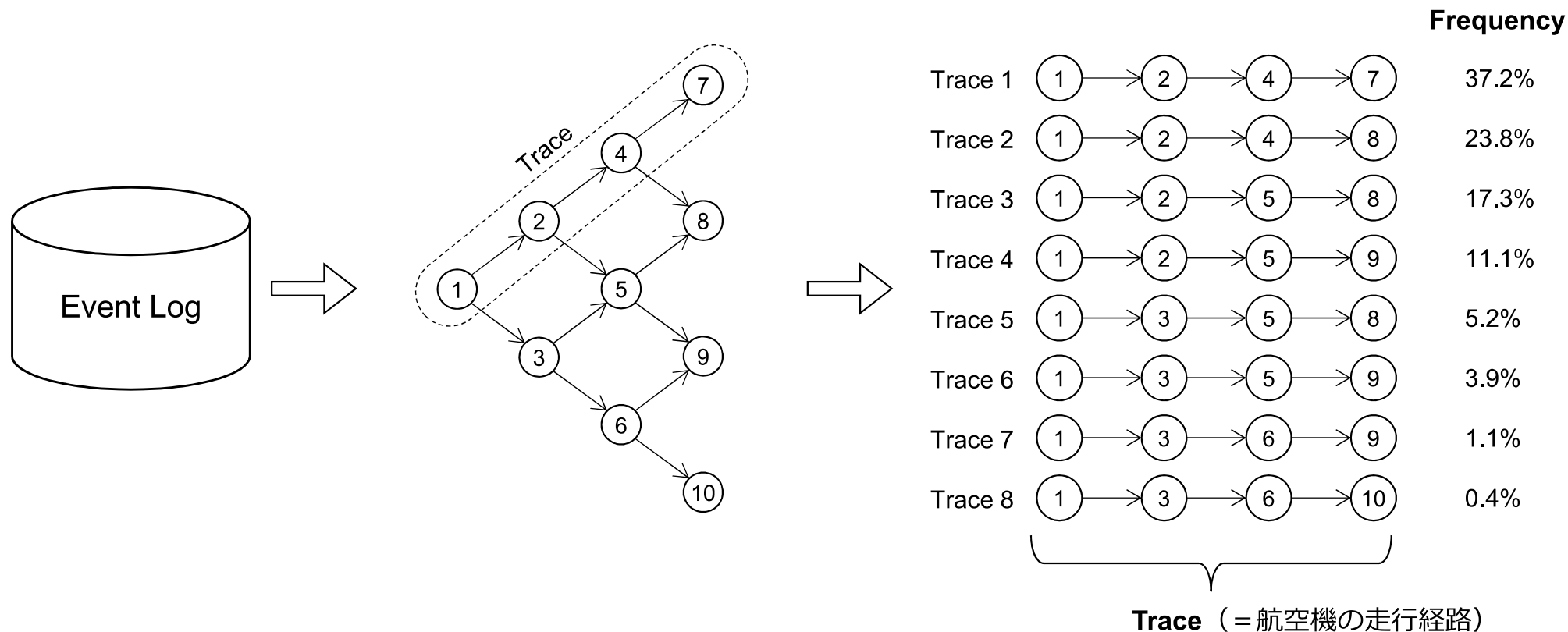
イベント分析では、各トレースの通過誘導路・滑走路をチェックしていきます。

Arr/Dep	Start/End	Events	Occurrence	Rate	Noise
Arrival	Start events	RWY	2658	99.81%	
		Other	5	0.19%	x
	End events	E-APRON	2052	77.06%	
		W-APRON	555	20.84%	
		Other	56	2.10%	x
Departure	Start events	E-APRON	2028	76.27%	
		W-APRON	557	20.95%	
		Other	74	2.78%	x
	End events	RWY	2652	99.74%	
		Other	7	0.26%	x

開始・終了イベントは必ず滑走路かエプロンでなければなりません。それ以外の場所で開始・終了している異常ケースがあります。これはCARATSデータに内在するノイズであり、排除することでモデル化のための綺麗なデータを作ることができます（=データクレンジング）。

12. トレース分析

トレース分析ではイベントログ中の状態遷移を調べ、可視化することができます。CARATSデータにおいては、全ての航空機が取った経路と発生頻度を列挙していきます。



通過した経路を調べ、その出現頻度を明らかにすることで、モデル化する際の航空機挙動の再現に活用できます。

13. トレース分析の結果

トレース分析により主要経路が分かる他、出現頻度の低いイレギュラーケースをノイズとして除去することができます。

出発経路（上位10）

#	Trace	Occurrence	Rate
1	<E-APRON, E1, RWY, END>	485	18.24%
2	<W-APRON, B2, B12, B11, W1, RWY, END>	362	13.61%
3	<E-APRON, A21, E2, RWY, END>	333	12.52%
4	<E-APRON, A22, A21, E2, RWY, END>	177	6.66%
5	<W-APRON, B42, B43, B5, B6, W9, RWY, END>	117	4.40%
6	<E-APRON, A21, A1, E1, RWY, END>	94	3.54%
7	<E-APRON, A1, A21, A22, A3, A4, A5, A6, A71, A72, A73, A74, E13, RWY, END>	93	3.50%
8	<E-APRON, E4, RWY, END>	92	3.46%
9	<E-APRON, A3, A4, A5, A6, A71, A72, A73, A74, E13, RWY, END>	90	3.38%
10	<E-APRON, A3, E4, RWY, END>	79	2.97%

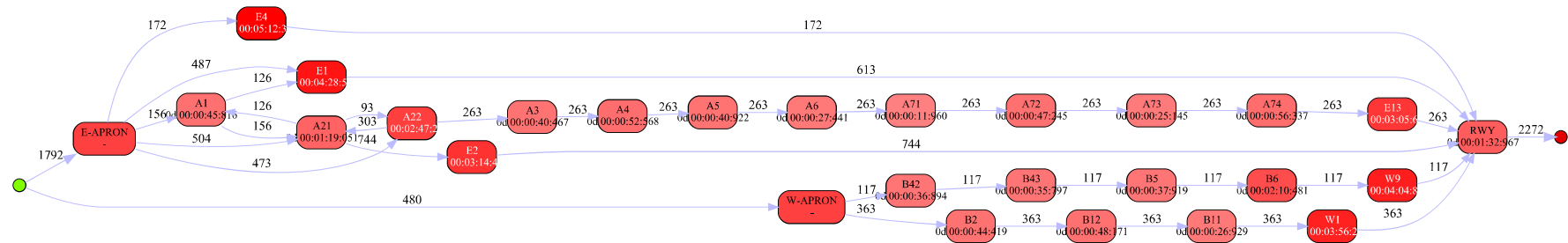
出発経路（イレギュラーケース）

#	Trace	Occurrence	Rate
1	<RWY, W4, B2, B12, B11, W1, E1, E-APRON, END>	1	0.04%
2	<RWY, W6, B43, END>	1	0.04%
3	<RWY, E1, END>	1	0.04%
4	<RWY, E8, A5, E7, RWY, E1, E-APRON, END>	1	0.04%
5	<E6, E-APRON, END>	1	0.04%
6	<RWY, E3, A21, END>	1	0.04%
7	<RWY, E10, A71, A6, A5, A4, A3, END>	1	0.04%
8	<RWY, W8, B5, END>	1	0.04%
9	<RWY, W4, END>	1	0.04%
10	<RWY, W2, B12, B11, W1, E2, A21, E-APRON, END>	1	0.04%

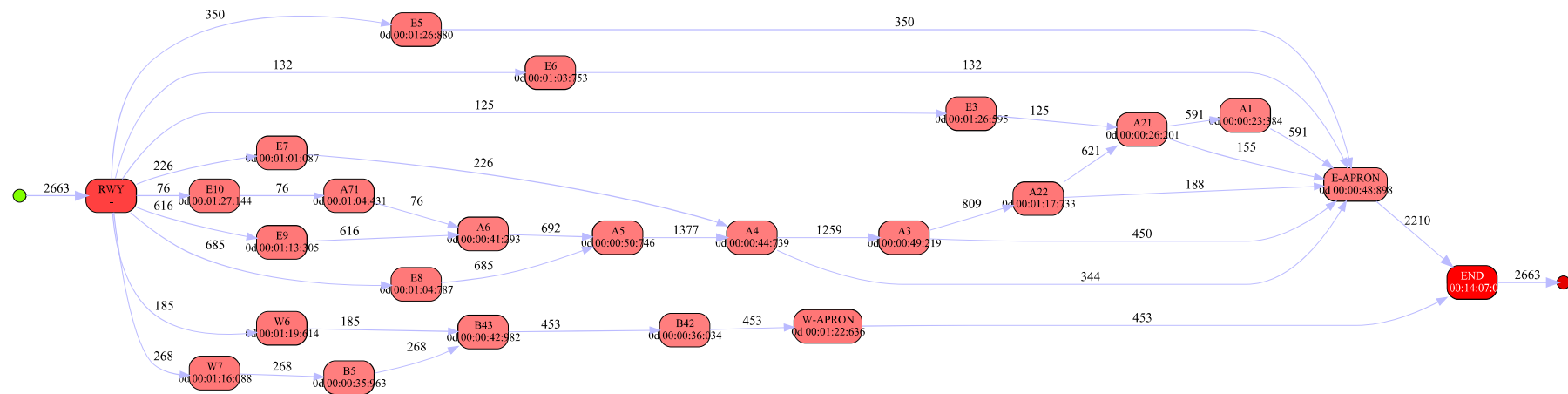
14. モデルの再現

プロセスマイニングは、トレースから状態遷移を再現することが可能です。遷移確率も分かるので、このままモデル化に利用することもできます。

出発機パターン

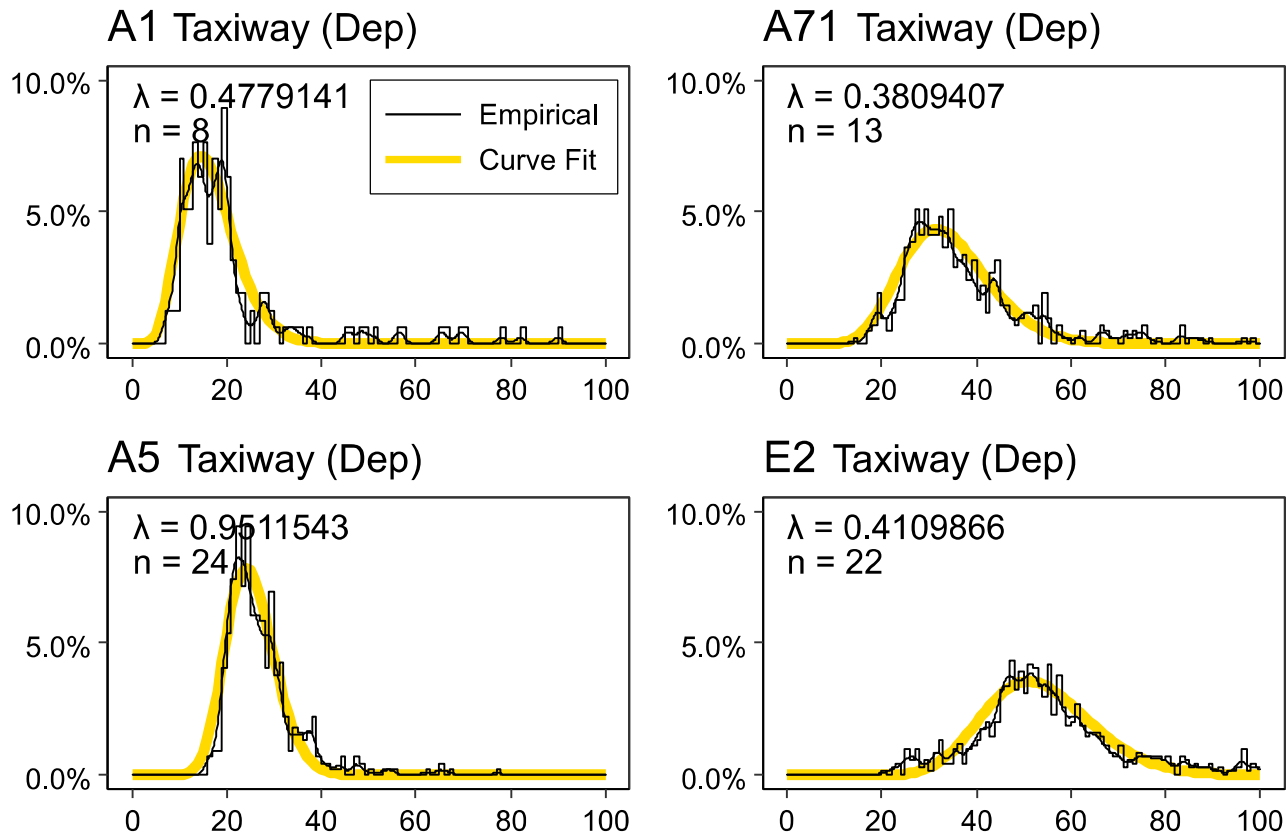


到着機パターン



15. 統計情報の抽出

ProMの機能により、各イベントの情報だけ切り取ることができます。各ノード間の移動情報だけを抽出し、統計情報としてモデル化に用います。

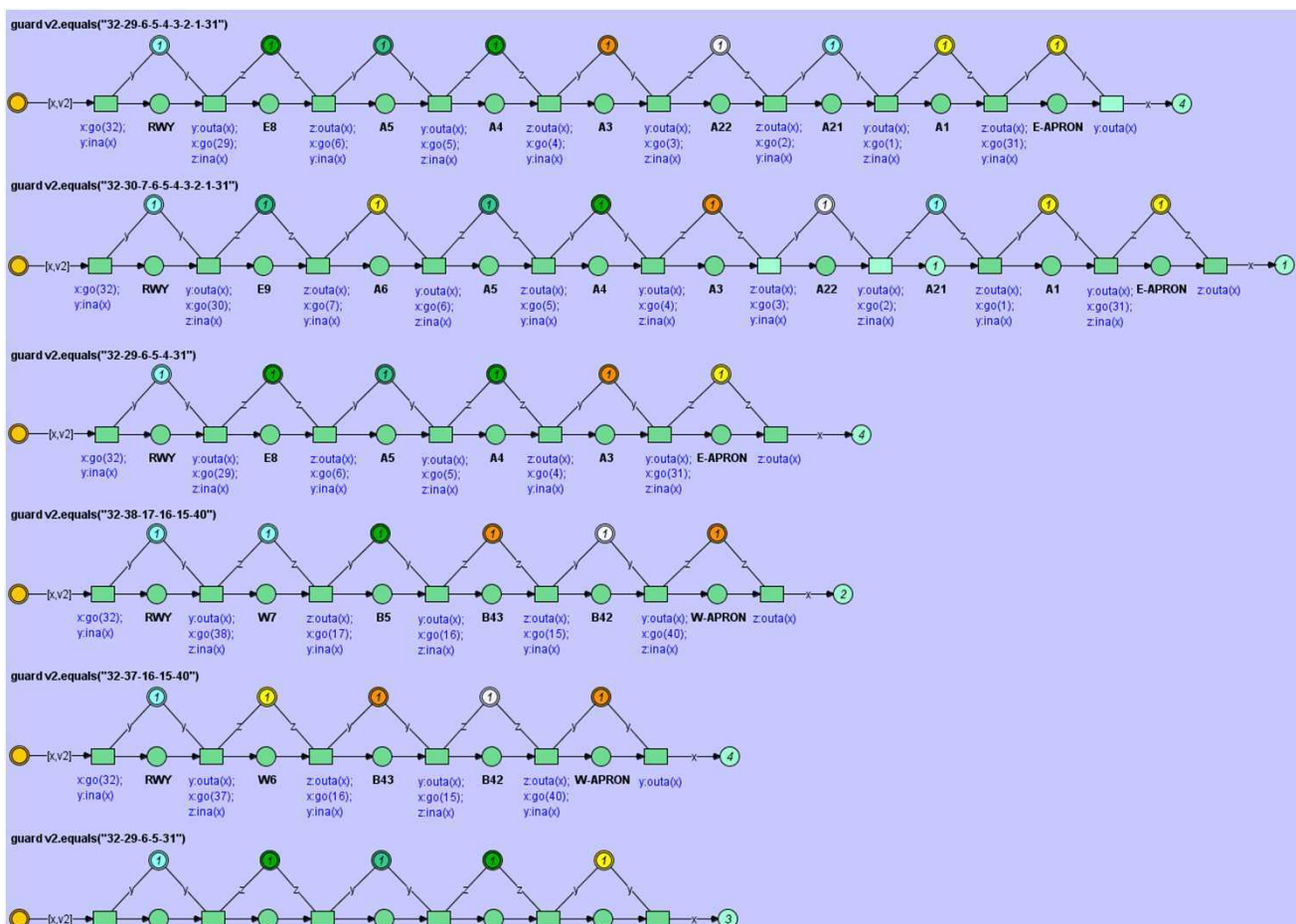


ここではさらに、各ノード間の移動情報を確率密度関数でフィッティングし、パラメータを抽出しています。

これにより、モデル化後は確率密度関数のみで航空機の移動を再現することができるようになり、**実データを持つ必要がなくなります。**

16. モデル化

プロセスマイニングのトレーズ解析結果及び統計情報を利用し、シミュレーションモデルを作成します。

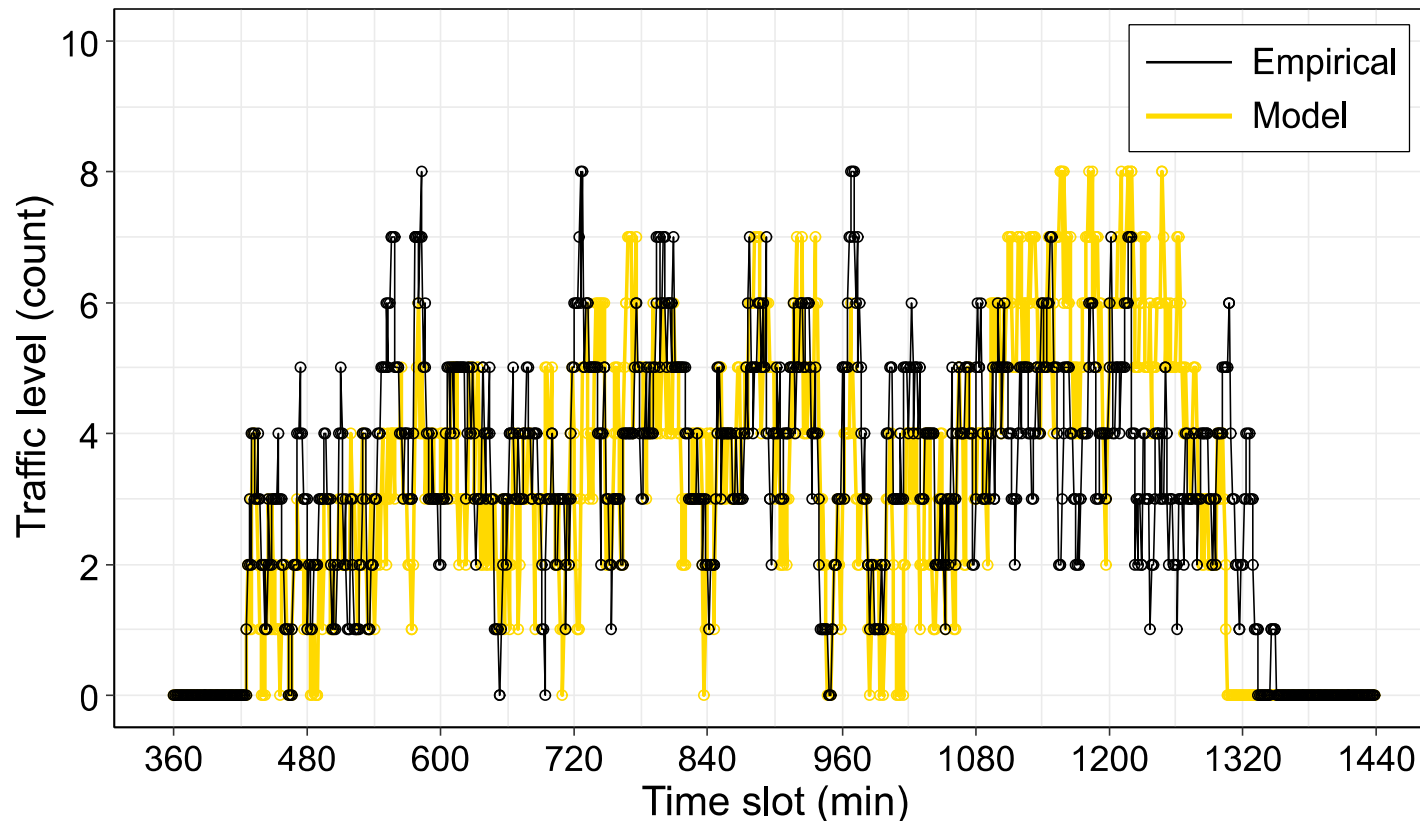


ここでは誘導路・滑走路の待ち行列を各トレースで共有させるようなモデルを作成しました。時間・場所をリソースとする競合モデルを構成することができます。

モデル化言語は、状態遷移が再現できるものであれば何を使っても良いです。

17. モデルの精度

CARATSデータとモデル出力の整合を確認します。ここでは交通量（= 空港面上に存在する航空機数）を60秒タイムスロットごとに比較してみました。



概ね良好な精度であり、追従性も良い。各場面における航空機位置をさらに細かく分析することで、滞留ポイントがどこにあるのかを把握することが可能です（=メゾスコピックの利点）。

18. まとめ

プロセスマイニングを応用し、交通流のメゾスコピックモデル化手法を紹介しました。

メゾスコピックモデルの利点

- データサイズを大幅に削減できる（CARATS空港面データの場合、約40分の1）
- 交通流の特徴を残しているため、フローやボトルネックの分析が可能
- 大まかな流れをパッと知りたい場合に適する

プロセスマイニングの利点

- イベントログのみでシステムの挙動解析が可能
- メゾスコピックモデル化に必要な統計データを効率的に抽出できる
- モデル化完了後も、挙動の検査に用いることができる

これからの課題

- 特に経路が決まっていない場合、交通の要所をどのように決定するか
- 空域データへの応用

ご清聴ありがとうございました。