5. Appendices

Appendix A. Delay Mechanism Pseudo-code

Table A1.: Pseudo-code for assigning delay propagation mechanism 2 - arrival backward propagation

Data: Darwin HSP
Result: Assign delay type 2 mechanism minutes
for *All Data* **do**
    if *Deviation from departure of primary train from station A = 0* **then**
        | Next iteration
    end
    Filter secondary train journeys on the same day
    Filter secondary train journeys that are arriving at station A
    Filter secondary train journeys that are scheduled to arrive at station A after primary train is
      scheduled to depart from station A
    Filter secondary train journeys that actually have scheduled arrival time at station A before
      primary train departs from station A
    for *Filtered journeys* **do**
        | Assign delay type 2 mechanism minutes
    end
end

Table A2.: Pseudo-code for assigning delay propagation mechanism 3 - arrival forward propagation

Data: Darwin HSP
Result: Assign delay type 3 mechanism minutes
for *All Data* **do**
    if *Deviation from arrival of primary train at station B = 0* **then**
        | Next iteration
    end
    Filter secondary train journeys on the same day
    Filter secondary train journeys that are arriving at station B
    Filter secondary train journeys that are scheduled to arrive at station B after primary train is
      scheduled to arrive at station B
    Filter secondary train journeys that actually have scheduled arrival time at station B before
      primary train arrives at station B
    for *Filtered journeys* **do**
        | Assign delay type 3 mechanism minutes
    end
end

Table A3.: Pseudo-code for assigning delay propagation mechanism 4 - departure backward propagation

Data: Darwin HSP
Result: Assign delay type 4 mechanism minutes
for *All Data* **do**
    if *Deviation from departure of primary train from station $A = 0$* **then**
        | Next iteration
    end
    Filter secondary train journeys on the same day
    Filter secondary train journeys that are departing from station A
    Filter secondary train journeys that are scheduled to depart from station A after primary train is scheduled to depart from station A
    Filter secondary train journeys that actually have scheduled departure time from station A before primary train departs from station A
    for *Filtered journeys* **do**
        | Assign delay type 4 mechanism minutes
    end
end

Table A4.: Pseudo-code for assigning delay propagation mechanism 5 - departure forward propagation

Data: Darwin HSP
Result: Assign delay type 5 mechanism minutes
for *All Data* **do**
    if *Deviation from departure of primary train from station $B = 0$* **then**
        | Next iteration
    end
    Filter secondary train journeys on the same day
    Filter secondary train journeys that are arriving at station B
    Filter secondary train journeys that are scheduled to depart from station B after primary train is scheduled to arrive at station B
    Filter secondary train journeys that actually have scheduled departure time from station B before primary train arrives at station B
    for *Filtered journeys* **do**
        | Assign delay type 5 mechanism minutes
    end
end

Appendix B. Data Preprocessing Pseudo-code

Table B1.: Pseudo-code for filling in missing actual train arrival time at current station

```
Data: Darwin HSP
Result: Filled in missing arrival time of current station
while Number of total null values change do
    for All Data do
        if TD_{C_{i-1}} ≠ null and TD_{C_i} ≠ null and i ≠ N then
            TA_{C_i} ← TD_{C_{i-1}} + Unique travel time between stations C_{i-1} and C_i
            if TA_{C_i} > TD_{C_i} then
                TA_{C_i} ← TD_{C_i} − Unique station dwell time of station C_i
                if TA_{C_i} < TD_{C_{i-1}} then
                    TA_{C_i} ← null
                end
            end
        end
        if TD_{C_{i-1}} ≠ null and TD_{C_i} = null and i = N then
            TA_{C_i} ← TD_{C_{i-1}} + Unique travel time between stations C_{i-1} and C_i
        end
        if TD_{C_{i-1}} = null and TD_{C_i} ≠ null and i ≠ N then
            TA_{C_i} ← TD_{C_i} − Unique station dwell time of station C_i
        end
    end
    Calculate total null values
end
```

Table B2.: Pseudo-code for filling in missing actual train departure time at current station

```
Data: Darwin HSP
Result: Filled in missing departure time of current station
while Number of total null values change do
    for All Data do
        if TA_{C_i} ≠ null and TA_{C_{i+1}} ≠ null and i ≠ 0 then
            TD_{C_i} ← TA_{C_{i+1}} − Unique travel time between stations C_i and C_{i+1}
            if TD_{C_i} < TA_{C_i} then
                TD_{C_i} ← TA_{C_i} + Unique station dwell time of station C_i
                if TD_{C_i} > TA_{C_{i+1}} then
                    TD_{C_i} ← null
                end
            end
        end
        if (TA_{C_{i+1}} ≠ null and i = 0) or (TA_{C_i} = null and TA_{C_{i+1}} ≠ null) then
            TD_{C_i} ← TA_{C_{i+1}} − Unique travel time between stations C_i and C_{i+1}
        end
        if TA_{C_i} ≠ null and TA_{C_{i+1}} = null and i ≠ 0 then
            TD_{C_i} ← TA_{C_i} + Unique station dwell time of station C_i
        end
    end
    Calculate total null values
end
```

Appendix C. Feature Engineering Process

Table C1.: Pre-processed data

Input Feature																				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	-1	Friday	0	-1	To Didcot Parkway	0	1.0	No	Yes	Feb	12	1	1	2	22.0	0	0	0	0	0
-3	-2	Friday	-3	-1	To Didcot Parkway	2	1.8	No	No	Feb	12	1	2	2	21.0	22	0	0	0	0
-5	-3	Friday	-2	-1	To Didcot Parkway	3	2.2	No	No	Feb	12	1	3	2	9.5	21	0	0	0	0
-7	-4	Friday	-2	-1	To Didcot Parkway	2	2.6	No	No	Feb	12	1	4	2	4.9	9	0	0	0	0
-9	-4	Friday	-2	0	To Didcot Parkway	3	2.0	No	No	Feb	12	1	5	2	9.9	5	0	0	0	0
-9	-5	Friday	0	-1	To Didcot Parkway	2	1.5	No	No	Feb	12	1	6	2	12.3	10	0	0	0	0
-9	-6	Friday	0	-1	To Didcot Parkway	1	2.3	No	No	Feb	12	1	7	2	10.9	12	0	0	0	0
-10	-6	Friday	-1	0	To Didcot Parkway	2	2.4	No	No	Feb	12	1	8	2	14.1	11	0	0	0	0
-10	4	Friday	0	10	To Didcot Parkway	11	3.4	No	No	Feb	12	1	9	2	17.2	15	0	0	0	0
1	22	Friday	11	18	To Didcot Parkway	10	3.7	No	No	Feb	12	1	10	2	12.4	18	0	0	0	11

Table C2.: Encoded pre-processed data

Input Feature																				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	-1	4	0	-1	1	0	1.0	0	1	1	12	1	1	2	22.0	0	0	0	0	0
-3	-2	4	-3	-1	1	2	1.8	0	0	1	12	1	2	2	21.0	22	0	0	0	0
-5	-3	4	-2	-1	1	3	2.2	0	0	1	12	1	3	2	9.5	21	0	0	0	0
-7	-4	4	-2	-1	1	2	2.6	0	0	1	12	1	4	2	4.9	9	0	0	0	0
-9	-4	4	-2	0	1	3	2.0	0	0	1	12	1	5	2	9.9	5	0	0	0	0
-9	-5	4	0	-1	1	2	1.5	0	0	1	12	1	6	2	12.3	10	0	0	0	0
-9	-6	4	0	-1	1	1	2.3	0	0	1	12	1	7	2	10.9	12	0	0	0	0
-10	-6	4	-1	0	1	2	2.4	0	0	1	12	1	8	2	14.1	11	0	0	0	0
-10	4	4	0	10	1	11	3.4	0	0	1	12	1	9	2	17.2	15	0	0	0	0
1	22	4	11	18	1	10	3.7	0	0	1	12	1	10	2	12.4	18	0	0	0	11

Table C3.: Encoded pre-processed data

| Input Feature | | | | | | | | | | | | | | |
Tue	Wed	Thu	Fri	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0

Table C4.: Input matrix

| Input Feature | | | | | | | | | | | | | | | | | |
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.36	-0.57
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.46	-0.60
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.54	-0.64
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.61	-0.67
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.68	-0.67
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.68	-0.70
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.68	-0.74
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.71	-0.74
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.71	-0.40
-0.48	-0.50	-0.51	-0.51	1.99	-0.30	-0.31	-0.30	-0.31	-0.30	-0.30	-0.29	-0.31	-0.30	-0.31	-0.28	-0.32	0.20

| Input Feature | | | | | | | | | | | | | | | | |
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
-0.35	-0.78	0.93	-0.77	-0.44	-0.30	3.30	-0.51	-1.62	-1.31	0.38	2.04	-1.09	-0.16	-0.15	-0.15	-0.23
-1.06	-0.78	0.93	0.24	0.28	-0.30	-0.30	-0.51	-1.62	-1.12	0.38	1.89	1.98	-0.16	-0.15	-0.15	-0.23
-0.82	-0.78	0.93	0.74	0.59	-0.30	-0.30	-0.51	-1.62	-0.93	0.38	0.24	1.84	-0.16	-0.15	-0.15	-0.23
-0.82	-0.78	0.93	0.24	0.99	-0.30	-0.30	-0.51	-1.62	-0.74	0.38	-0.43	0.17	-0.16	-0.15	-0.15	-0.23
-0.82	-0.52	0.93	0.74	0.47	-0.30	-0.30	-0.51	-1.62	-0.55	0.38	0.29	-0.39	-0.16	-0.15	-0.15	-0.23
-0.35	-0.78	0.93	0.24	0.01	-0.30	-0.30	-0.51	-1.62	-0.35	0.38	0.64	0.31	-0.16	-0.15	-0.15	-0.23
-0.35	-0.78	0.93	-0.27	0.69	-0.30	-0.30	-0.51	-1.62	-0.16	0.38	0.44	0.59	-0.16	-0.15	-0.15	-0.23
-0.59	-0.52	0.93	0.24	0.82	-0.30	-0.30	-0.51	-1.62	0.03	0.38	0.90	0.45	-0.16	-0.15	-0.15	-0.23
-0.35	2.04	0.93	4.77	1.70	-0.30	-0.30	-0.51	-1.62	0.22	0.38	1.35	1.00	-0.16	-0.15	-0.15	-0.23
2.24	4.10	0.93	4.27	1.92	-0.30	-0.30	-0.51	-1.62	0.41	0.38	0.65	1.42	-0.16	-0.15	-0.15	2.02

Appendix D. DNN vs XGBoost Results
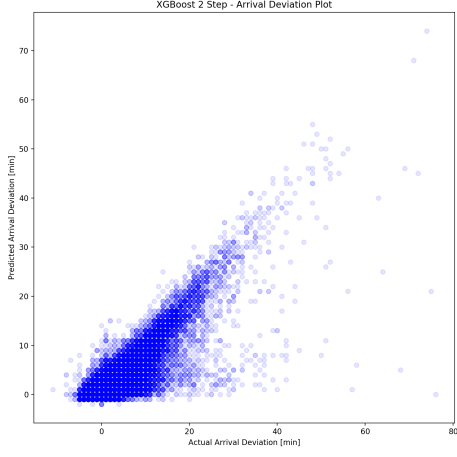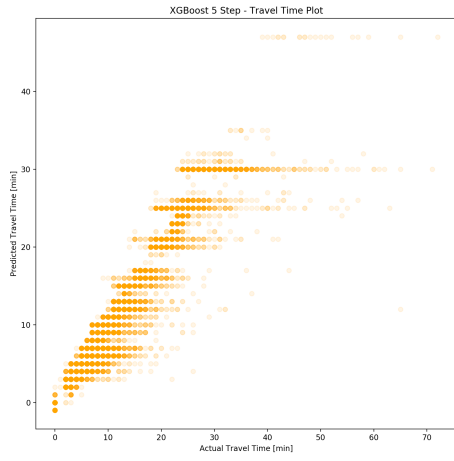

(a) 2-step DNN deviation from arrival
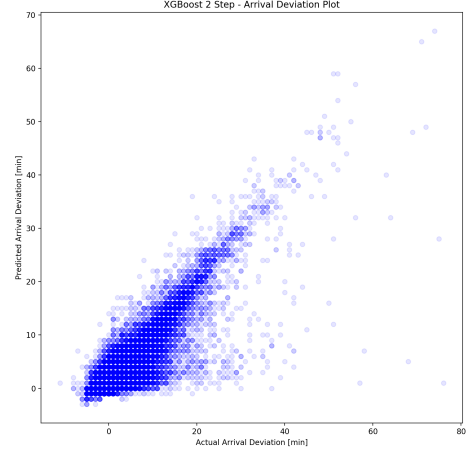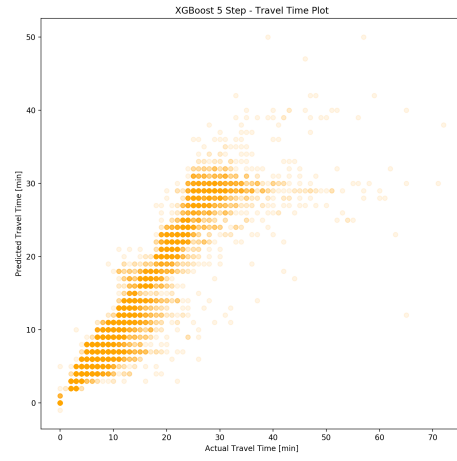

(b) 2-step XGBoost deviation from arrival
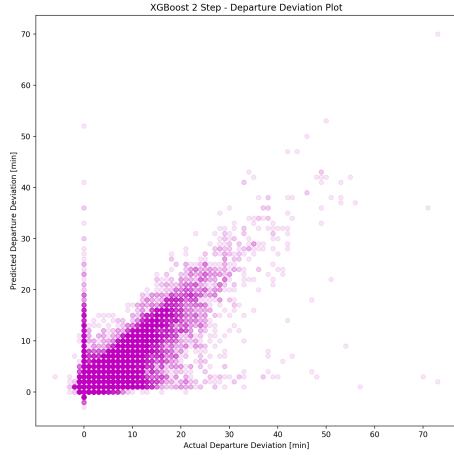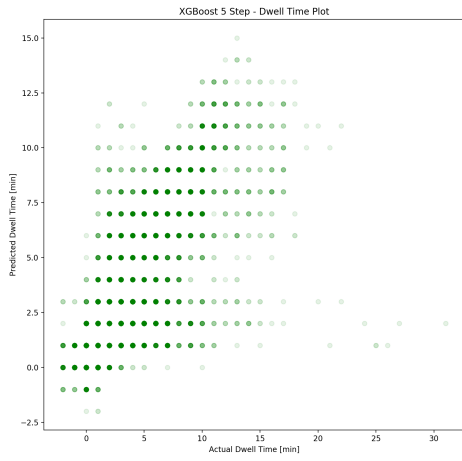

(c) 2-step DNN deviation from departure
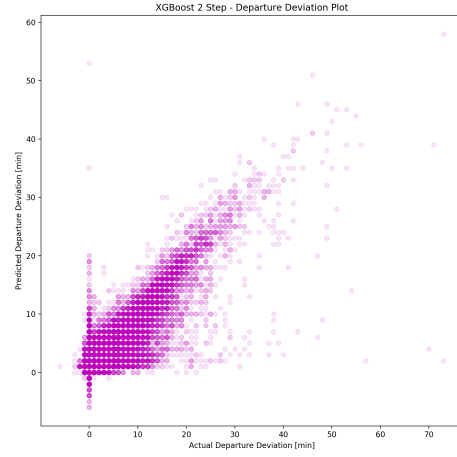
(d) 2-step XGBoost deviation from departure
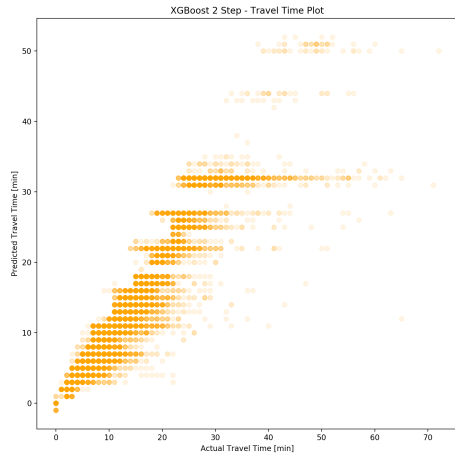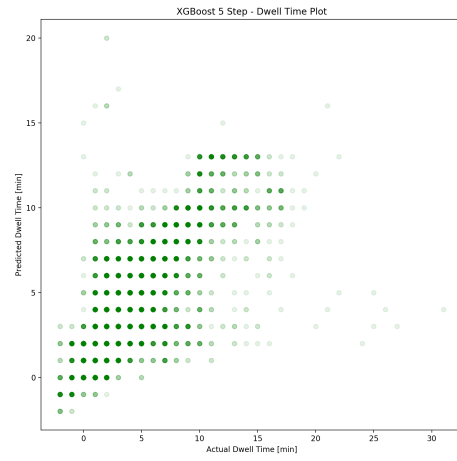
(e) 2-step DNN travel time
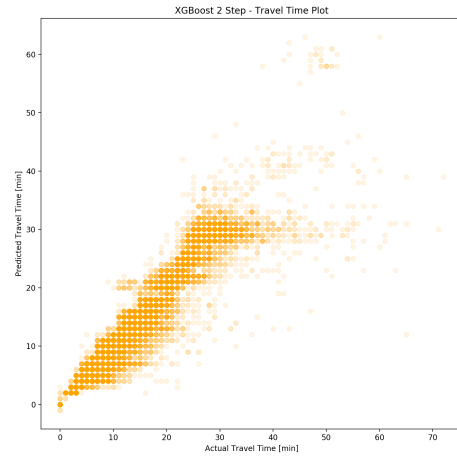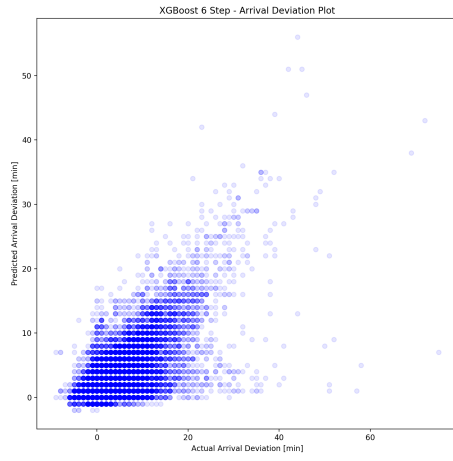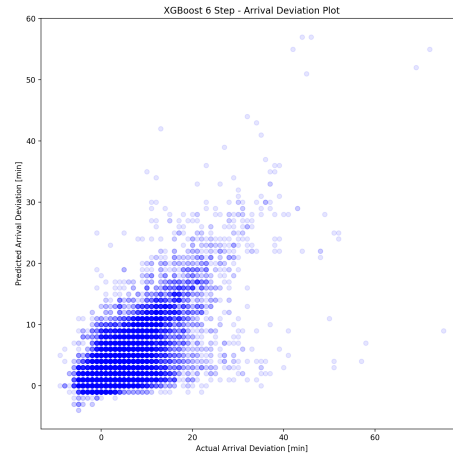

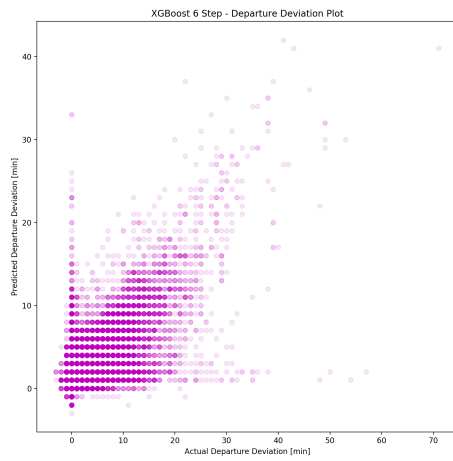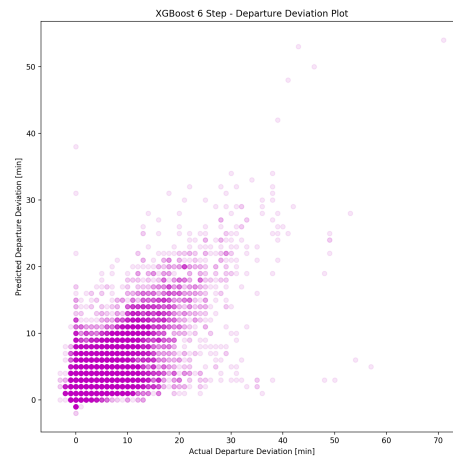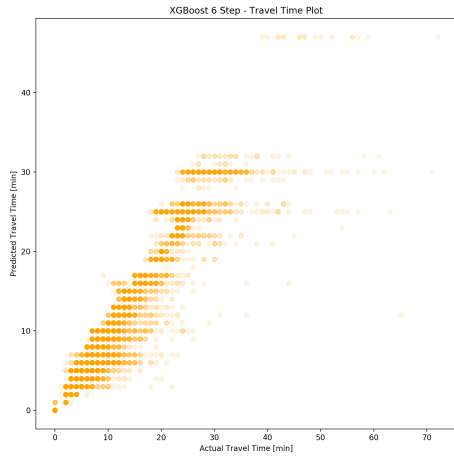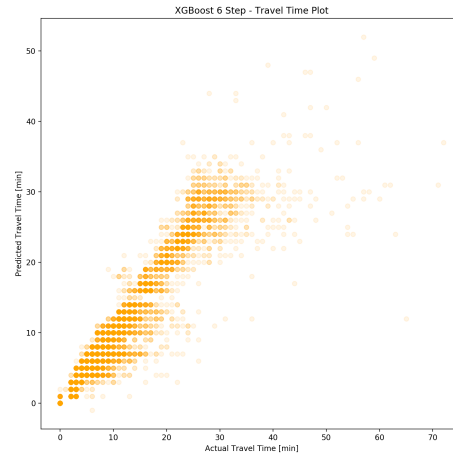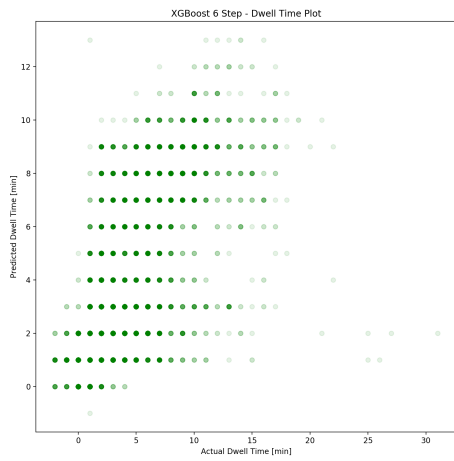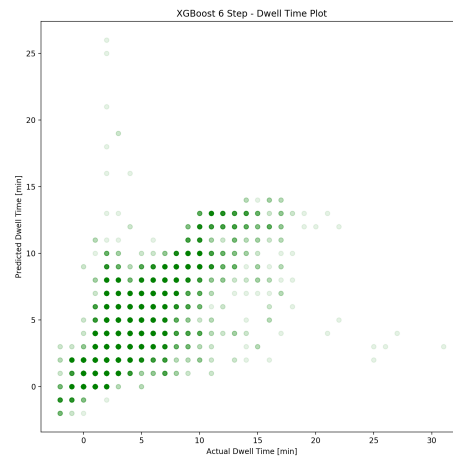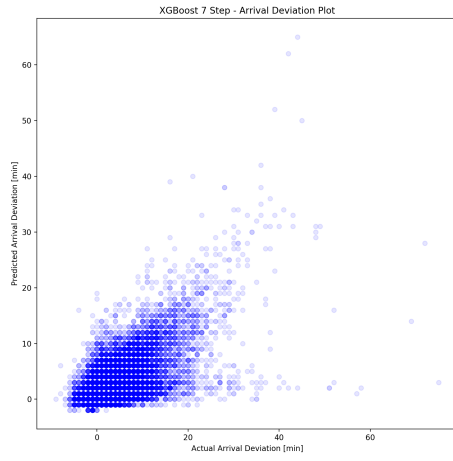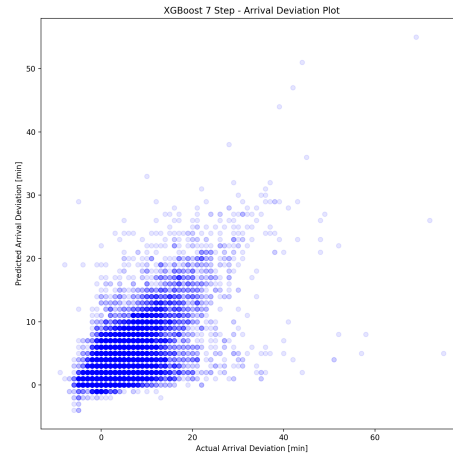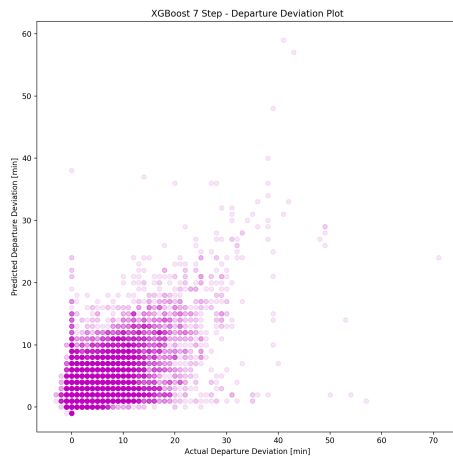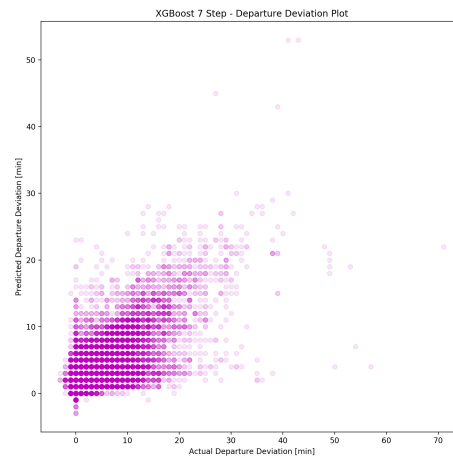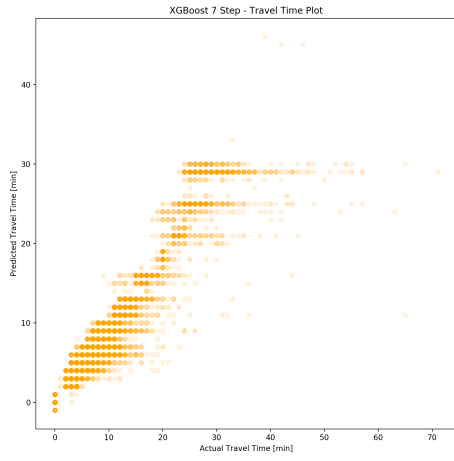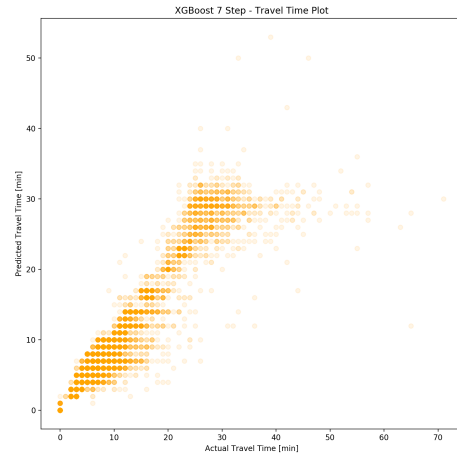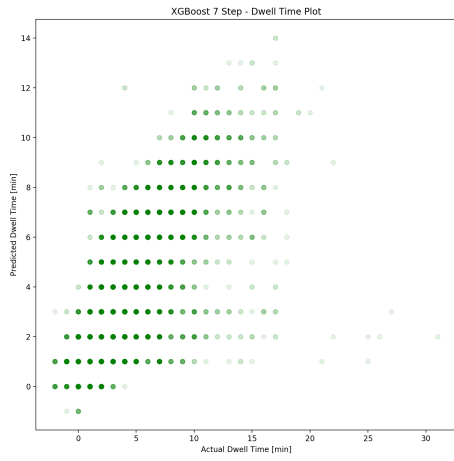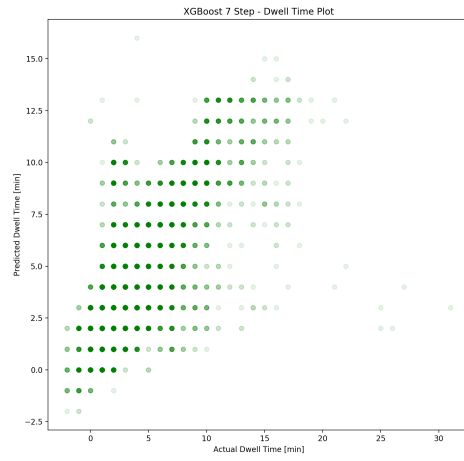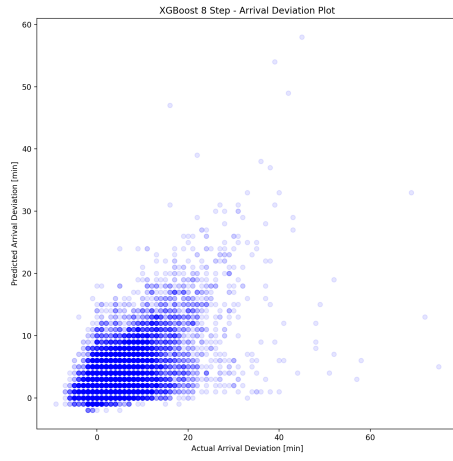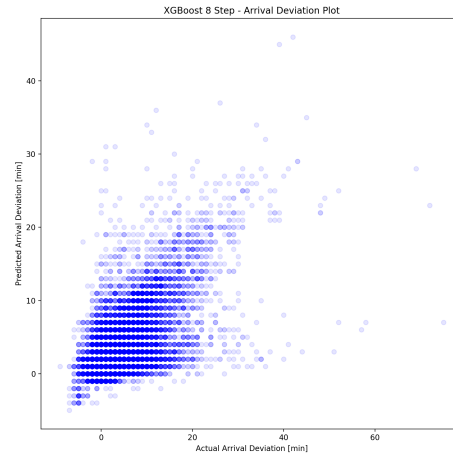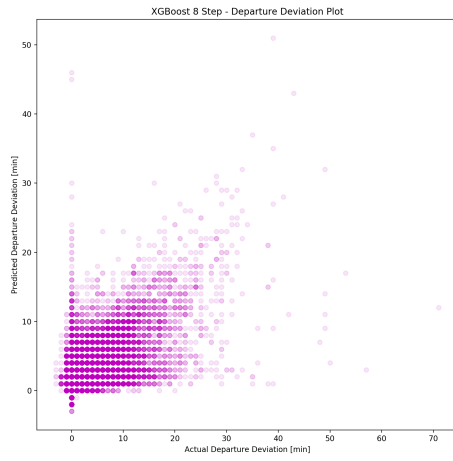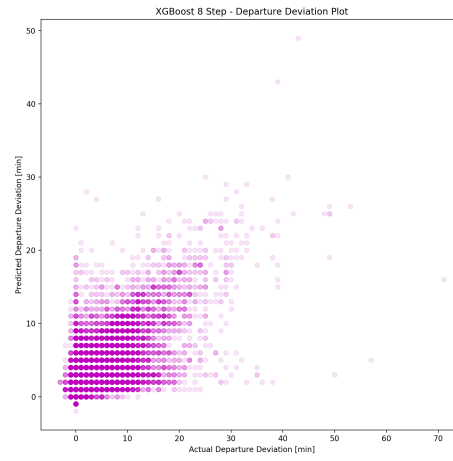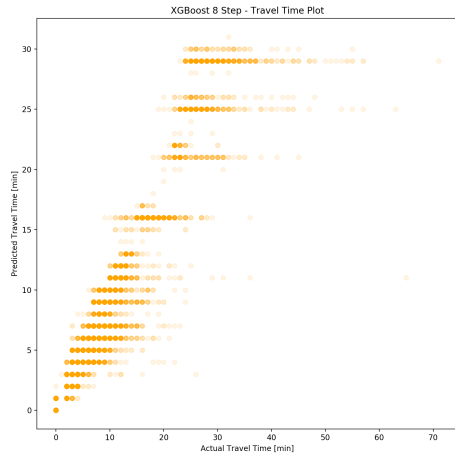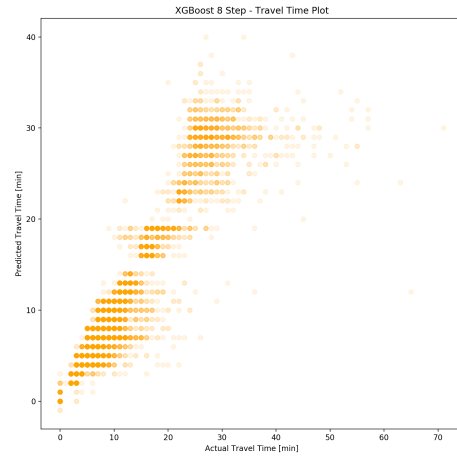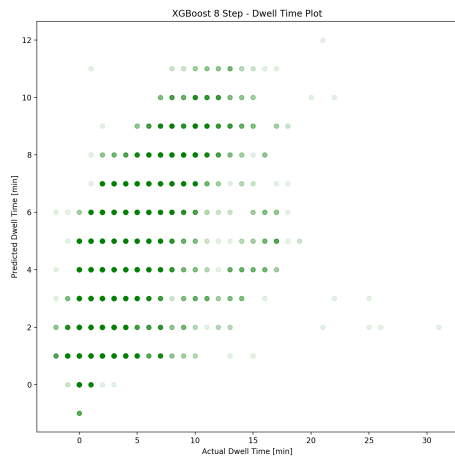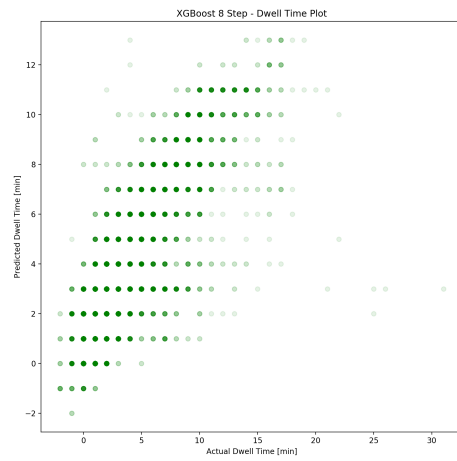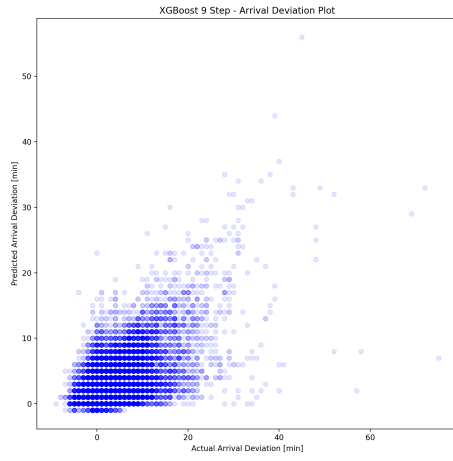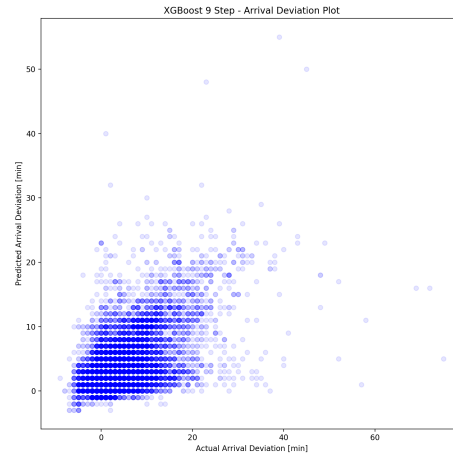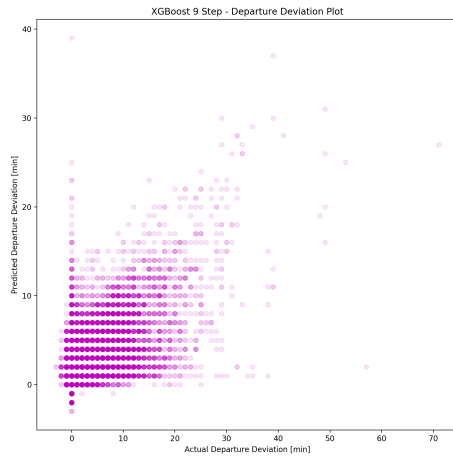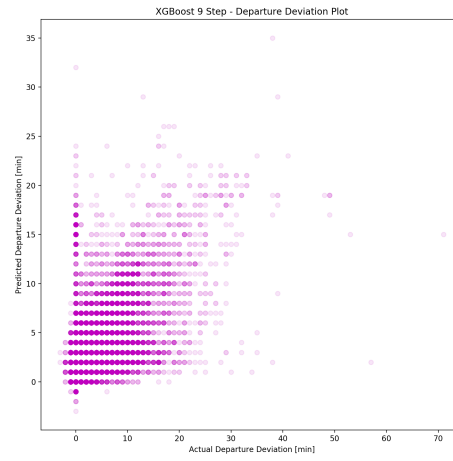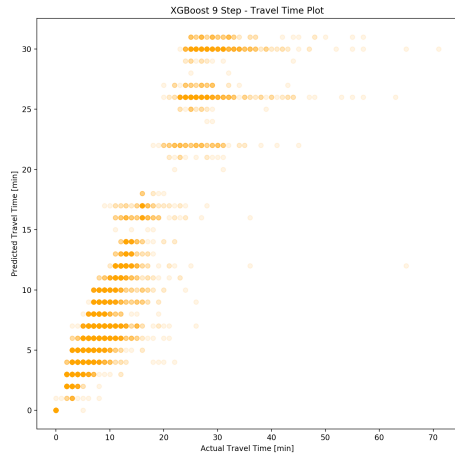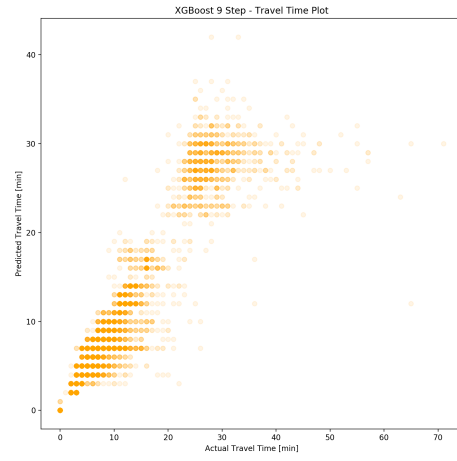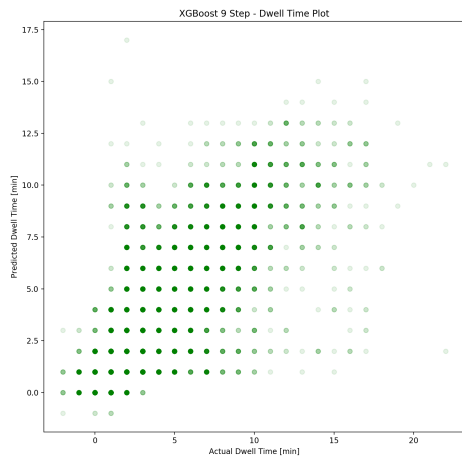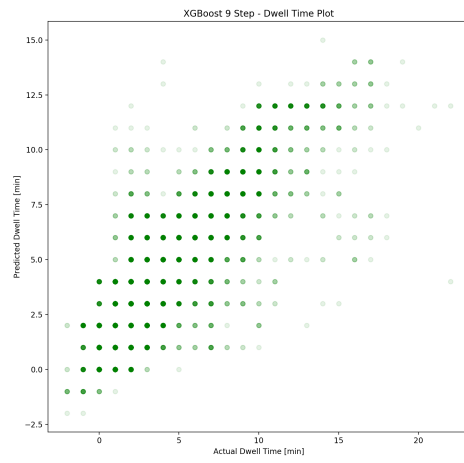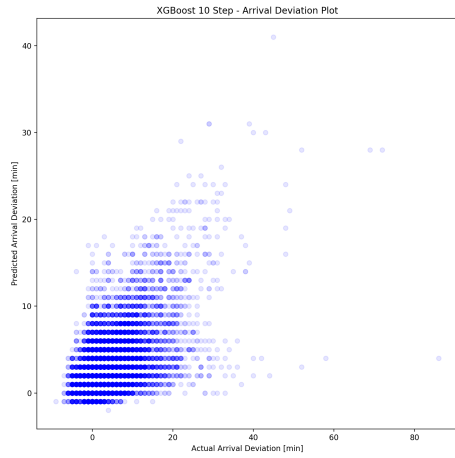(f) 2-step XGBoost travel time
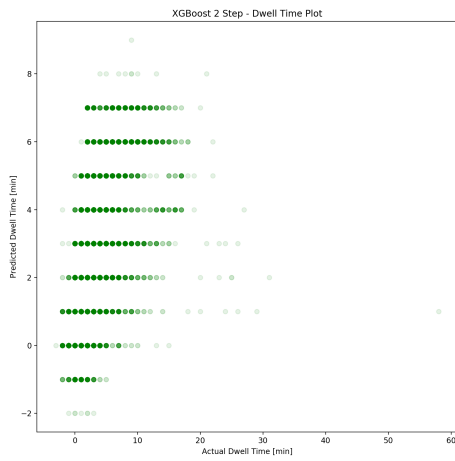
(g) 2-step DNN dwell time
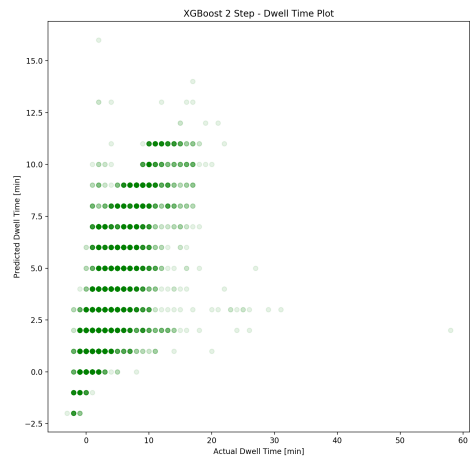
(h) 2-step XGBoost dwell time
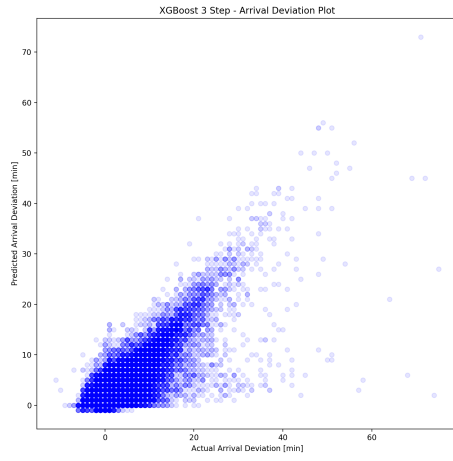
(i) 3-step DNN deviation from arrival
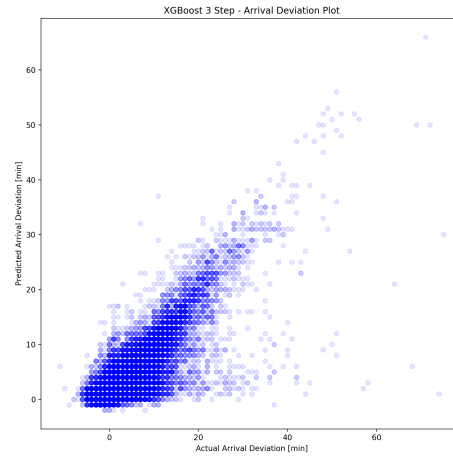
(j) 3-step XGBoost deviation from arrival
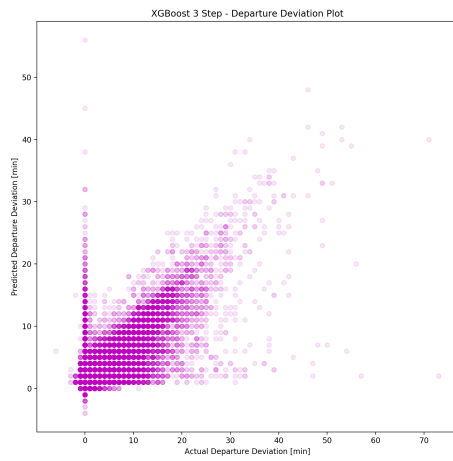
(k) 3-step DNN deviation from departure
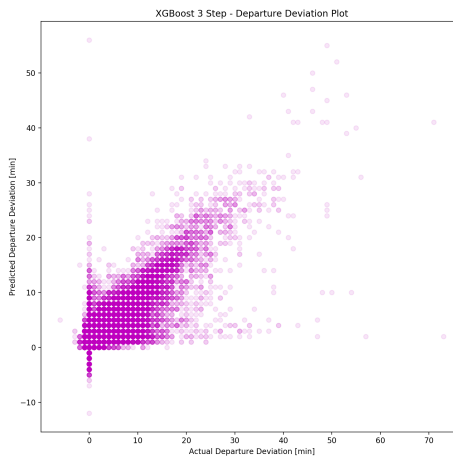
(l) 3-step XGBoost deviation from departure
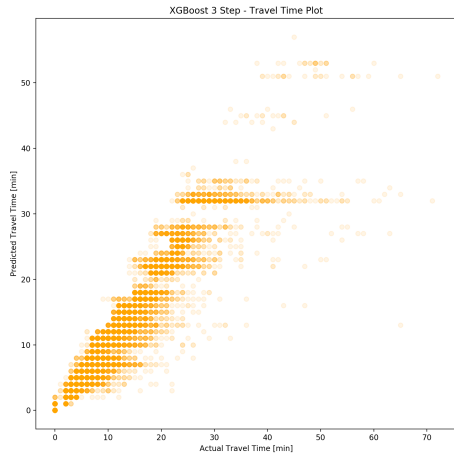
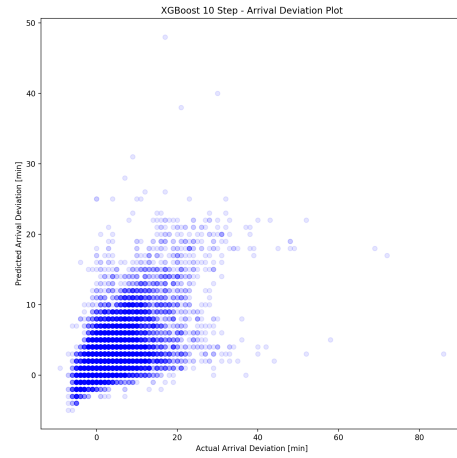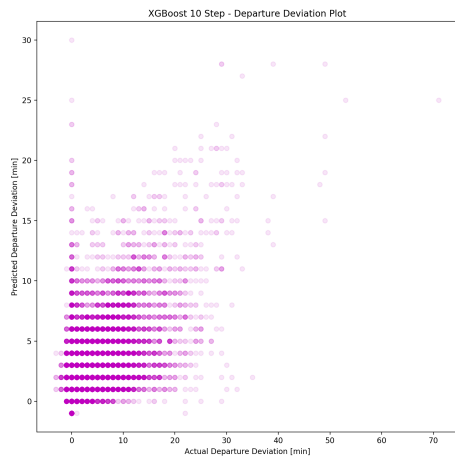(m) 3-step DNN travel time
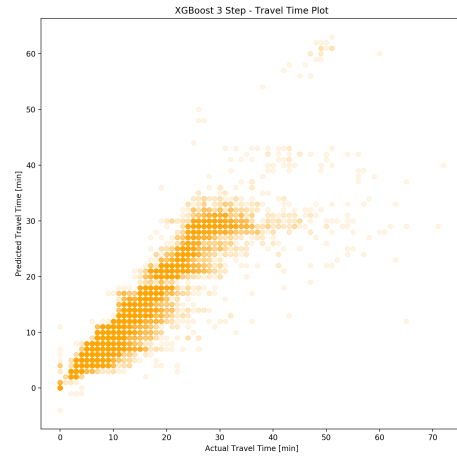
(n) 3-step XGBoost travel time
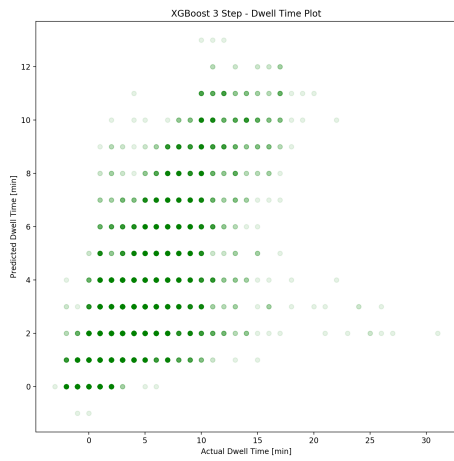
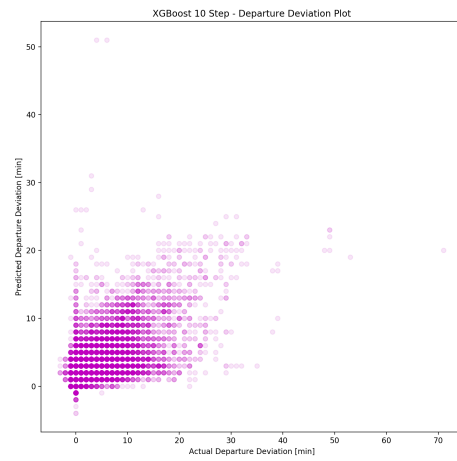(o) 3-step DNN dwell time
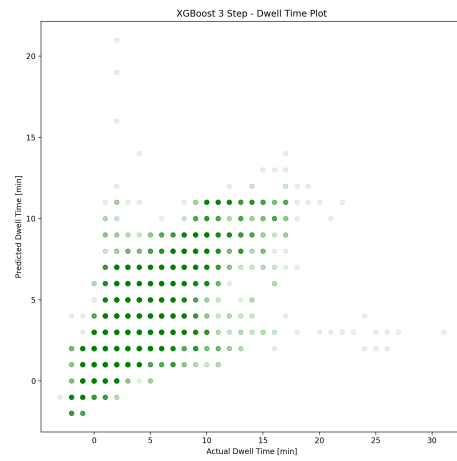
(p) 3-step XGBoost dwell time
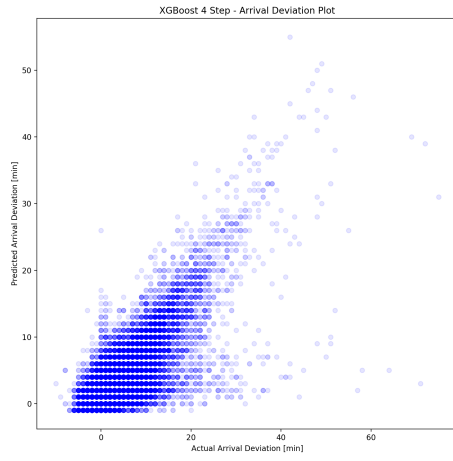
(q) 4-step DNN deviation from arrival
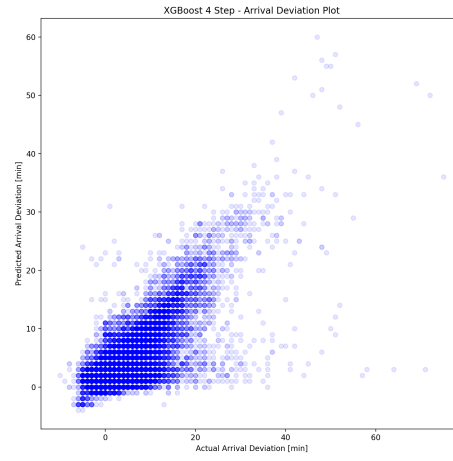

(r) 4-step XGBoost deviation from arrival
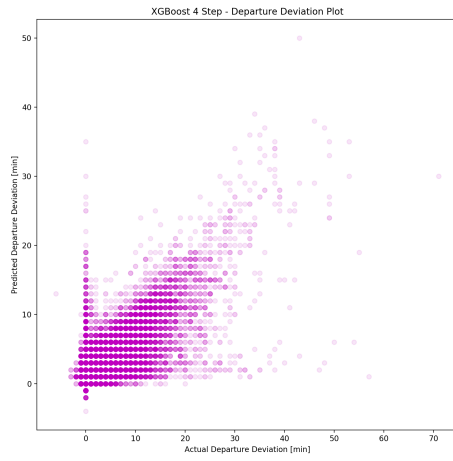

(s) 4-step DNN deviation from departure
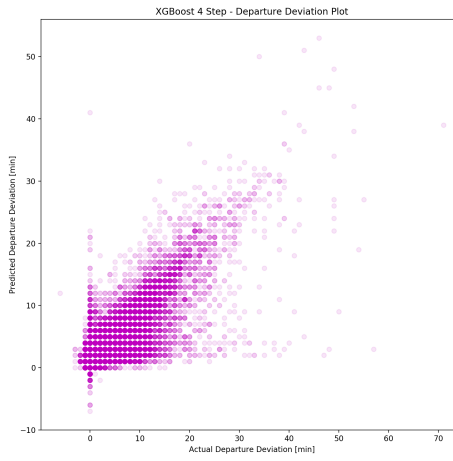

(t) 4-Step XGBoost Deviation from Departure
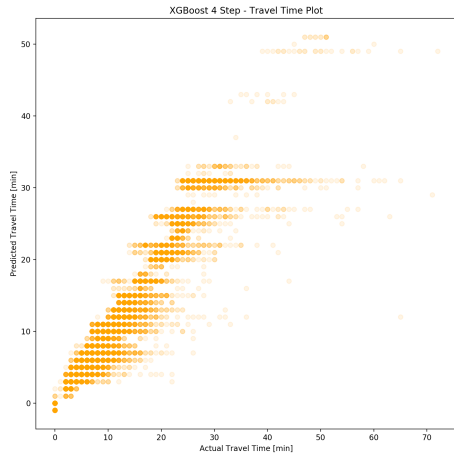
(u) 4-step DNN travel time
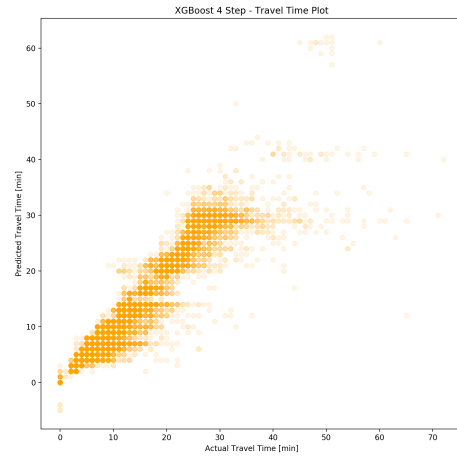
(v) 4-step XGBoost travel time
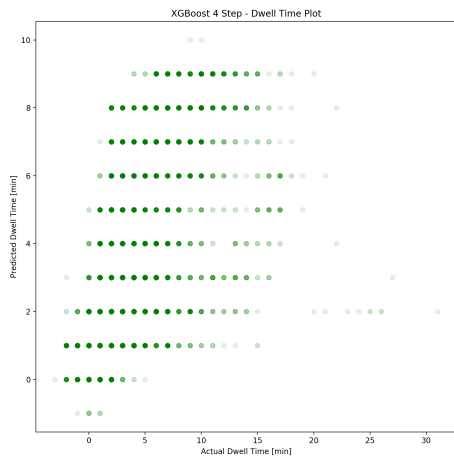
(w) 4-step DNN dwell time
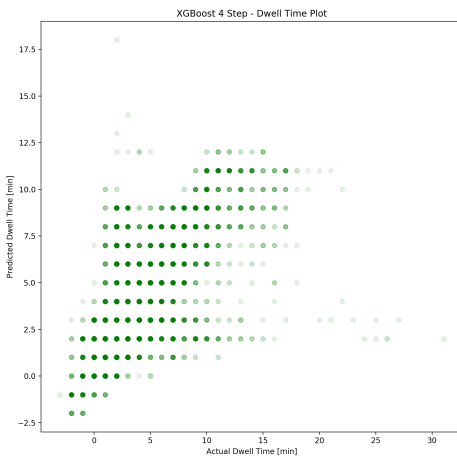
(x) 4-step XGBoost dwell time
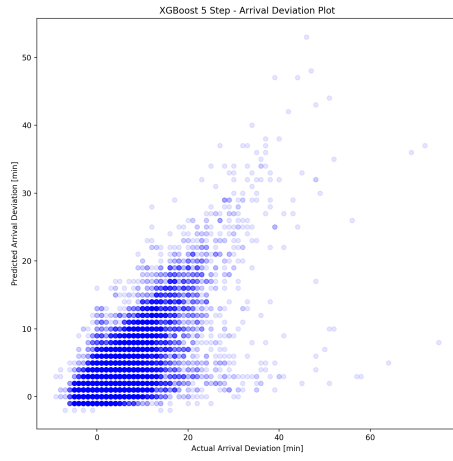
(y) 5-step DNN deviation from arrival
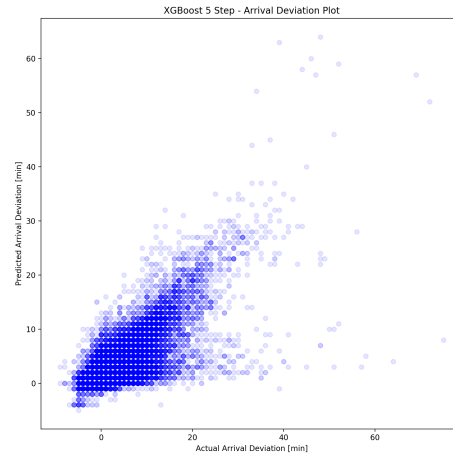
(z) 5-step XGBoost deviation from arrival
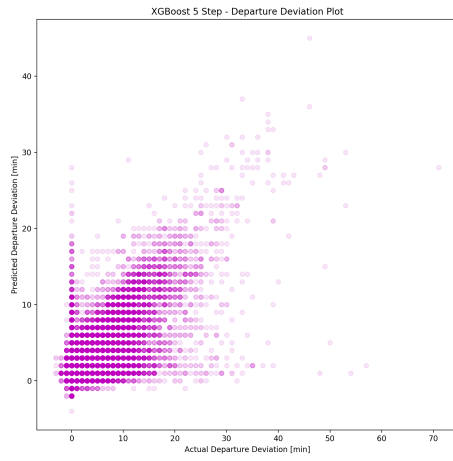
(aa) 5-step DNN deviation from departure
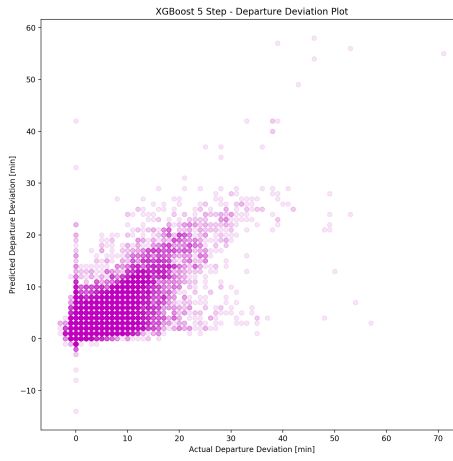

(ab) 5-step XGBoost deviation from departure

(ac) 5-step DNN travel time

(ad) 5-step XGBoost travel time



(ae) 5-step DNN dwell time

(af) 5-step XGBoost dwell time

(ag) 6-step DNN deviation from arrival

(ah) 6-step XGBoost deviation from arrival

(ai) 6-step DNN deviation from departure

(aj) 6-step XGBoost deviation from departure

(ak) 6-step DNN travel time

(al) 6-step XGBoost travel time

(am) 6-step DNN dwell time

(an) 6-step XGBoost dwell time

(ao) 7-step DNN deviation from arrival

(ap) 7-step XGBoost deviation from arrival

(aq) 7-step DNN deviation from departure

(ar) 7-step XGBoost deviation from departure

(as) 7-step DNN travel time

(at) 7-step XGBoost travel time

(au) 7-step DNN dwell time

(av) 7-step XGBoost dwell time

(aw) 8-step DNN deviation from arrival

(ax) 8-step XGBoost deviation from arrival

(ay) 8-step DNN deviation from departure

(az) 8-step XGBoost deviation from departure

(ba) 8-step DNN travel time

(bb) 8-step XGBoost travel time

(bc) 8-step DNN dwell time

(bd) 8-step XGBoost dwell time

(be) 9-step DNN deviation from arrival

(bf) 9-step XGBoost deviation from arrival

(bg) 9-step DNN deviation from departure

(bh) 9-step XGBoost deviation from departure

(bi) 9-step DNN travel time

(bj) 9-step XGBoost travel time

(bk) 9-step DNN dwell time

(bl) 9-step XGBoost dwell time

(bm) 10-step DNN deviation from arrival

(bn) 10-step XGBoost deviation from arrival

(bo) 10-step DNN deviation from departure

(bp) 10-step XGBoost deviation from departure
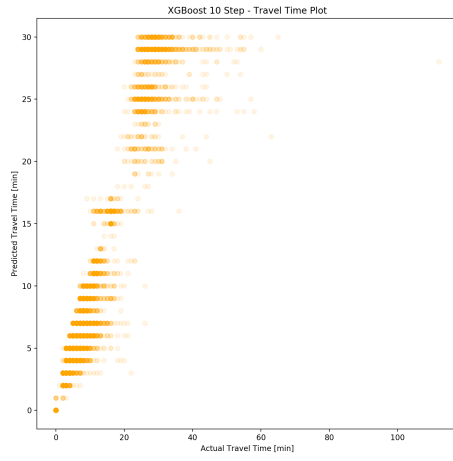
(bq) 10-step DNN travel time
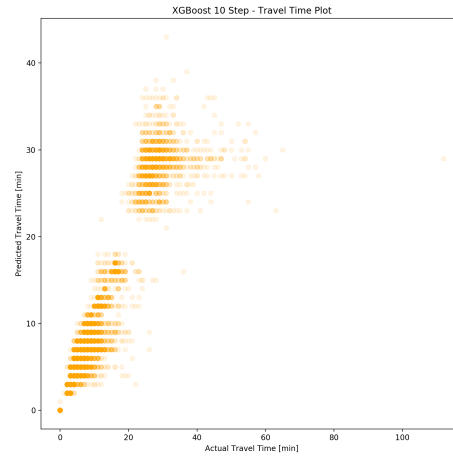
(br) 10-step XGBoost travel time
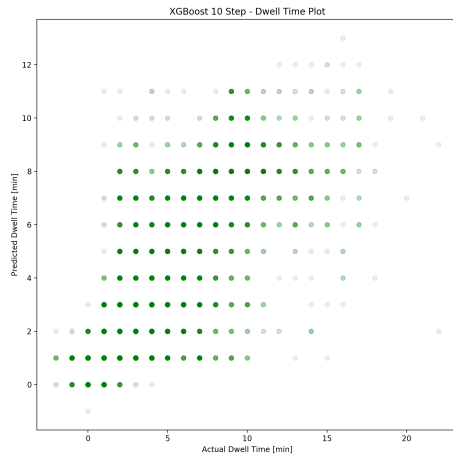
(bs) 10-step DNN dwell time
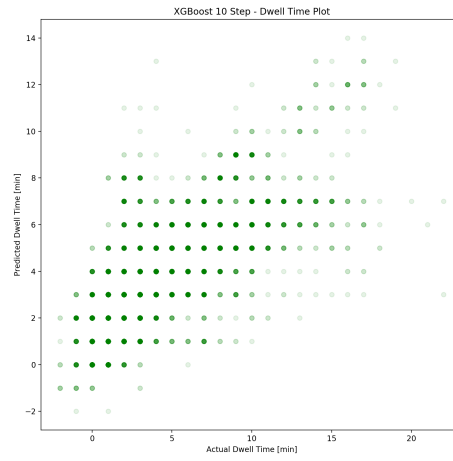
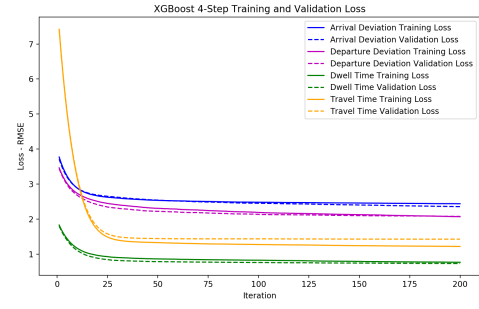(bt) 10-step XGBoost dwell time

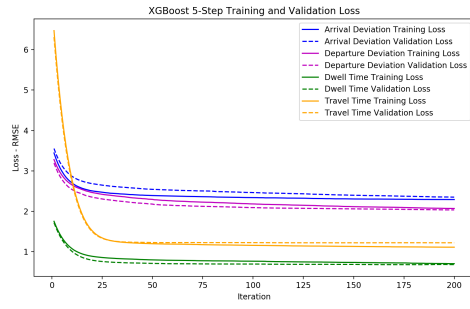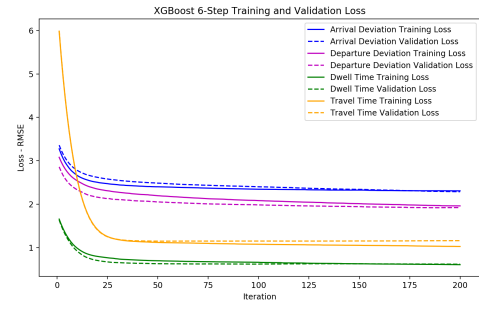## Appendix E. DNN Training and Validation Loss
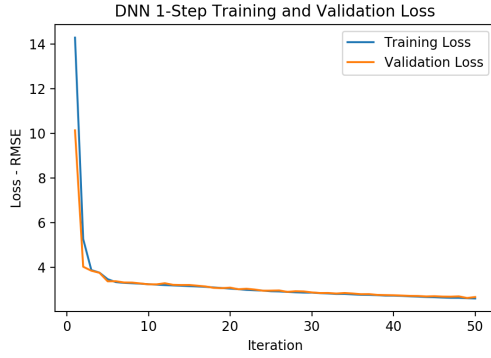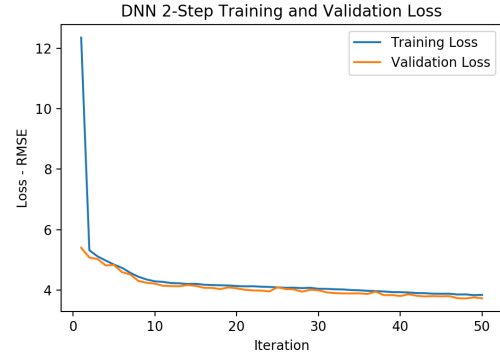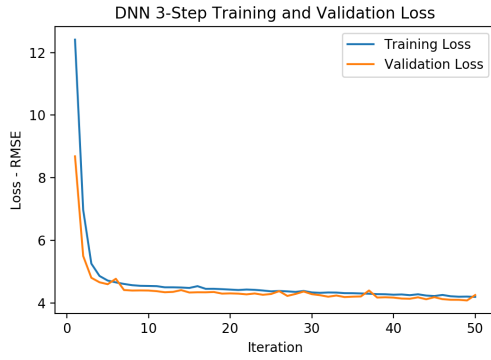
(a) 1-step
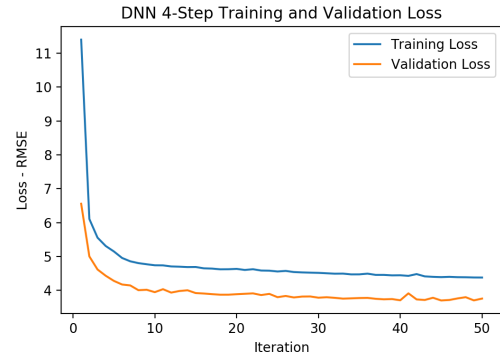
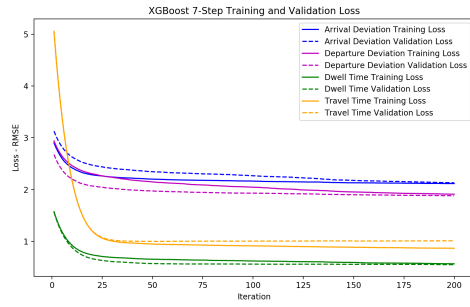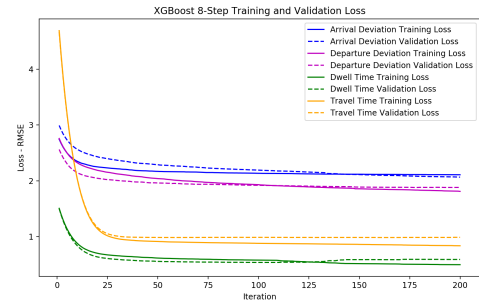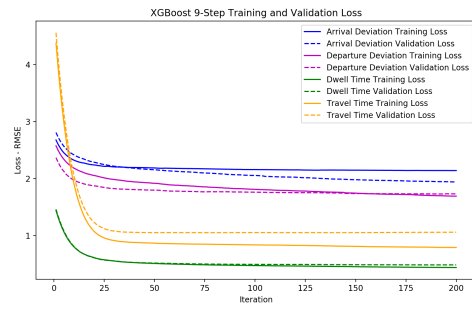(b) 2-step
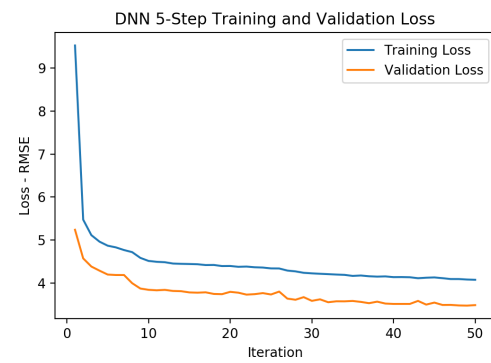
(c) 3-step
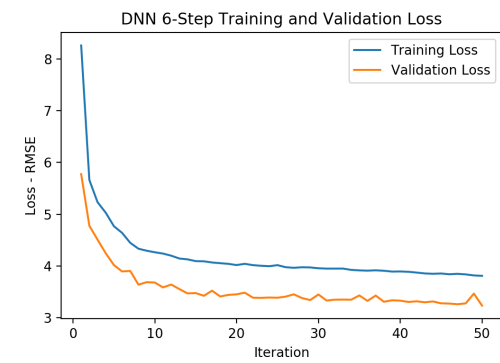
(d) 4-step
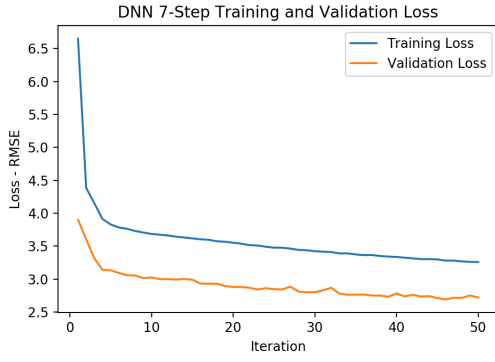
(e) 5-step
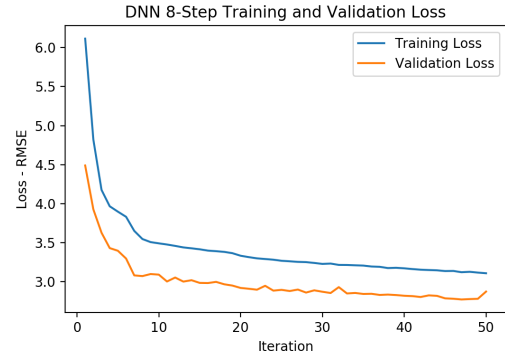
(f) 6-step
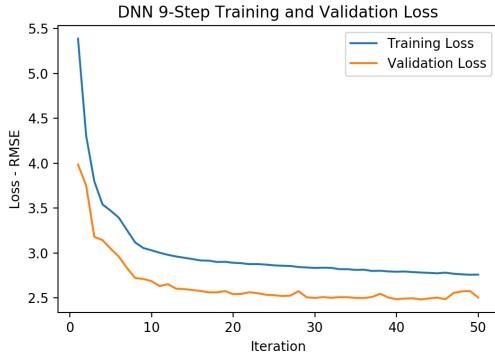
(g) 7-step
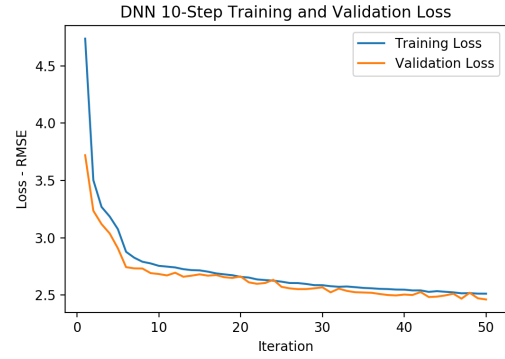
(h) 8-step

(i) 9-step

(j) 10-step

Appendix F. XGBoost Training and Validation Loss

(a) 1-step

(b) 2-step

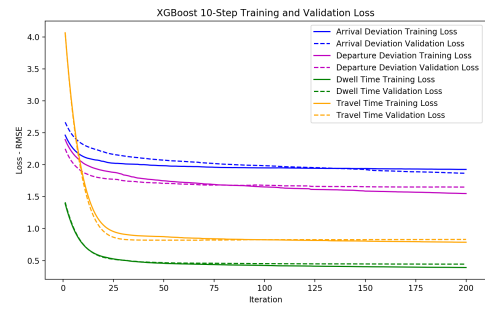(c) 3-step

(d) 4-step

(e) 5-step

(f) 6-step

(g) 7-step

(h) 8-step

(i) 9-step

(j) 10-step