

BWC数据源质量对比

整体性 | 及时性 | 准确性 | 总结



制作: Fuller





目录



一、数据源



二、整体性



三、及时性



四、准确性

01

02

03

04

目录



一、数据源



航班数据源

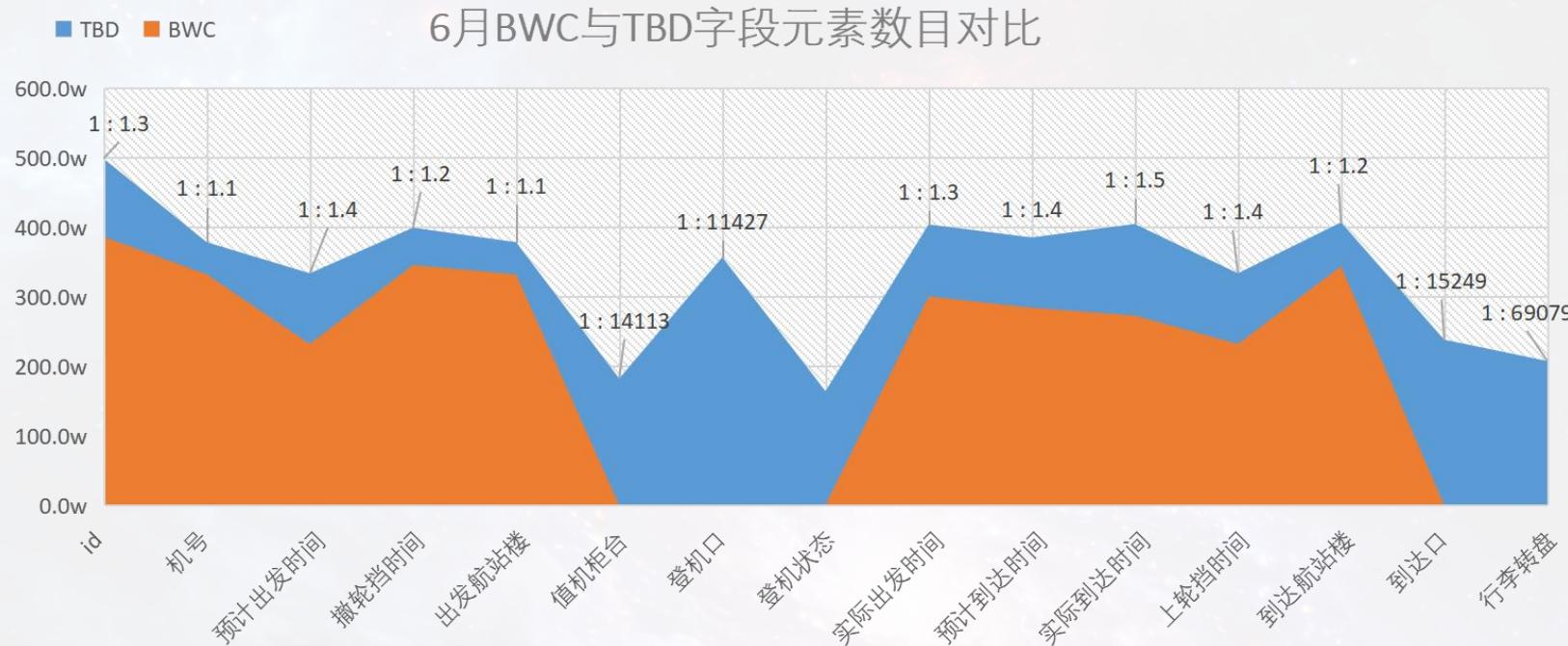
- BWC航班
- TBD航班
- 备注：取2077-06月

日志数据源

- BWC更新日志
- TBD更新日志



二、整体性-6月全部记录



- BWC与TBD的总记录数比例在 **1 : 1.3**。
- 在机号、预飞时间、推出时间等字段的比例在**1 : 1.2**左右。但在值机柜台、登机口、登机状态、到达口、行李等旅客服务信息方面，BWC信息较少，相比TBD**差距明显**。

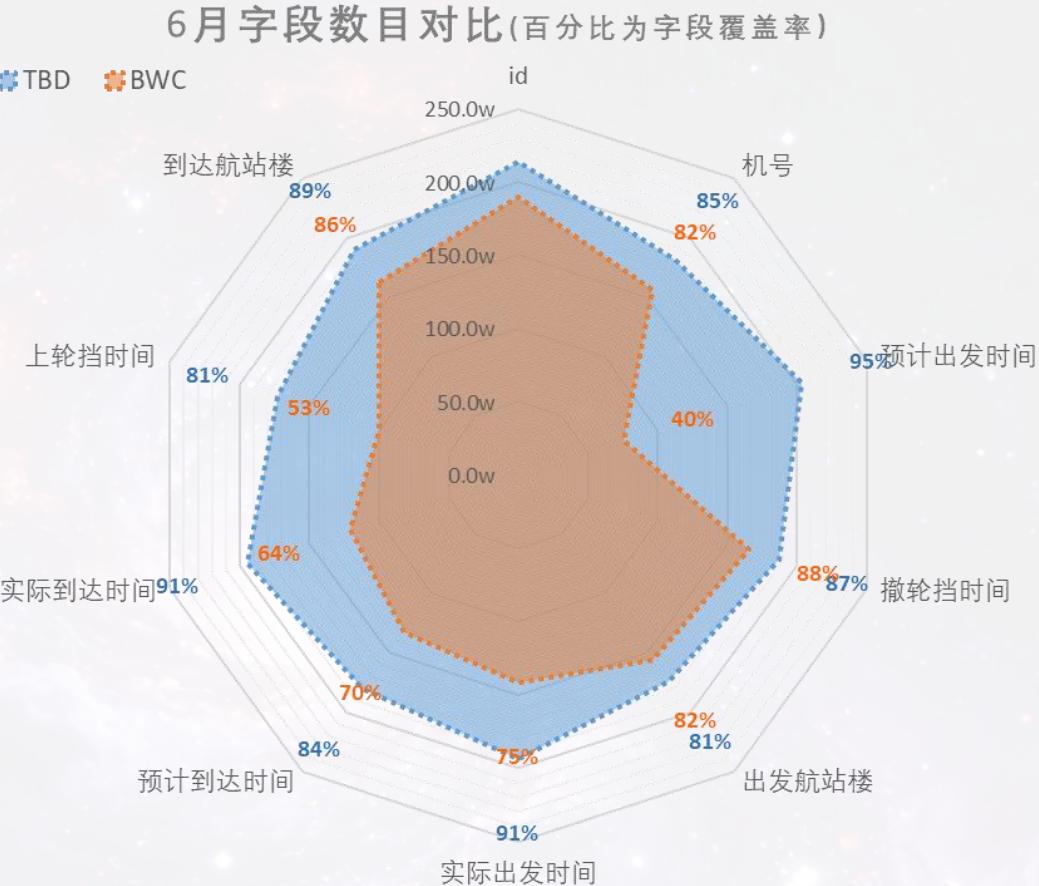


二、整体性-6月执飞记录

- 6月正常航班的记录，BWC与TBD的记录比例在 1 : 1.1。
- TBD各字段的**覆盖率均在80%以上**，而BWC的仅有机号、预飞、推出时间三个字段的覆盖率超80%。

备注：筛选执飞航班；

字段覆盖率 = 有效值数目 / 记录总数



(Tips: 本报告数据已脱敏处理，无实际意义，仅作报告结构与样式展示。)



三、及时性-首更时间对比

- 逻辑说明：对比BWC和TBD 更新日志中各个字段首次更新的时间，计算差值分析两处数据源的更新时间早晚；其中**负值表示BWC更新更早，反之则TBD更新更早，0表示同时更新。**
- 从右表中可以明显看出首次更新时间差值的分布。如实降时间、机号等字段，两者50%以上的记录更新时间相同，而如到达航站、出发航站、预降时间等字段，**TBD有更早更新的数据渠道。**
- 右图通过分析1万7千条取消航班记录，TBD普遍比BWC提前更新，即**TBD获取航班取消信息更加及时。**

及时差 / min	实际到达时间	实际出发时间	机号	到达航站楼	出发航站楼	预计到达时间	上轮挡时间	撤轮挡时间
<-60	2099	146	2549	2888	1800	53373	209	95
-60~-45	134	7	10	6	5	4176	12	1
-45~-30	106	12	6	3	1	5672	10	16
-30~-15	113	15	16	10	4	13555	28	11
-15~-5	94	8	13	0	0	12020	59	38
-5~0	400	22	7	2	1	48171	187	49
0	385194	285043	627049	95752	87732	16915	325647	377245
0~5	37556	40723	14674	632	474	30244	51832	38605
5~15	35665	36695	9878	540	604	20842	54254	41463
15~30	30021	28158	8891	518	394	15041	36983	32244
30~45	29942	26677	8258	420	281	13462	30948	30262
45~60	29488	27299	7310	279	303	14439	31415	29283
>=60	154337	165420	81368	187568	187784	215606	153664	175869
及状态对航班比新班	10089	9453	642	571	430	621	3393	
	0	0~5	5~15	15~30	30~45	45~60	>=60	

(Tips: 本报告数据已脱敏处理，无实际意义，仅作报告结构与样式展示。)

四、准确性-预降时间误差



- 右侧三图预降时间准确性分析：实际时间-预计时间，按误差在30min内为准确， TBD的准确率在98%， BWC的准确率在99%。按误差在5min内为准确， TBD与BWC的预计准确率均在70%；
- 二者在预计时间的准确占比上无明显差异，但TBD在数据量上占绝对优势；

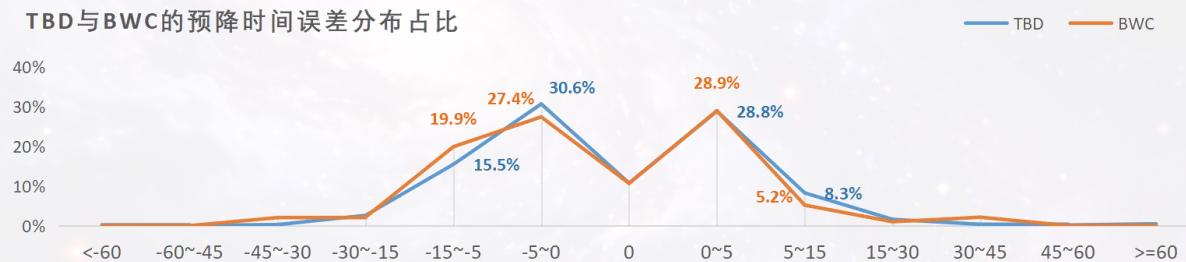
TBD预降时间误差分布



BWC预降时间误差分布



TBD与BWC的预降时间误差分布占比



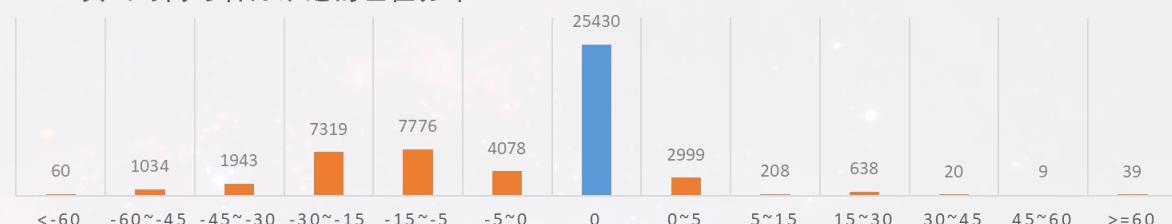
(Tips: 本报告数据已脱敏处理，无实际意义，仅作报告结构与样式展示。)

四、准确性-BWC对比官方渠道

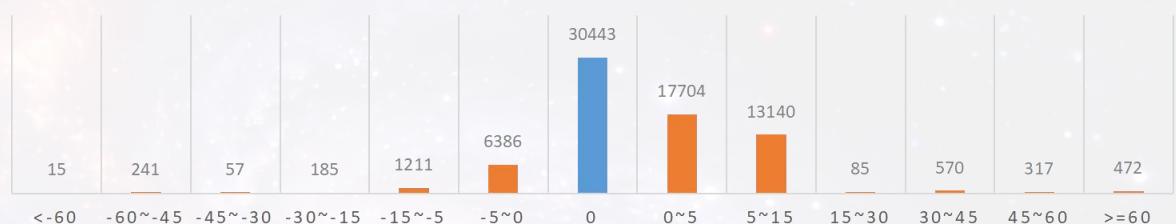
BWC实际起落时间与**官方渠道结果**对比



BWC实飞时间与官方渠道的差值分布

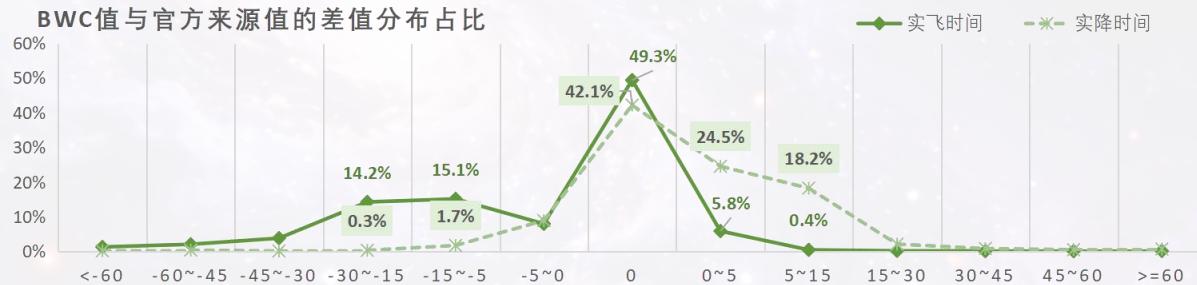


BWC实降时间与官方渠道的差值分布



- 上图是BWC的实际起落时间与官方渠道值的对比结果，可以看出**BWC中50%以上**的实际起落时间与官方来源数值是不一致的，实降时间的不一致占比更是接近60%。
- 右侧三图延伸分析了BWC渠道数值与官方渠道数值的差值分布，**负值表示BWC的时间比官方延后，正值表示BWC的时间比官方要早**；
- 从右图可以看出，BWC实飞时间里约37%的记录比官方实飞时间延后约0~30min，而BWC实降时间中约45%的记录比官方到达时间提早0~30min。

BWC值与官方来源值的差值分布占比



(Tips: 本报告数据已脱敏处理，无实际意义，仅作报告结构与样式展示。)



- BWC与TBD的航班总记录数比例在 1 : 1.3，即TBD的记录数量**比BWC多50%**；
- 正常执飞的国际客运航班记录中，**TBD各字段的覆盖率均在90%以上**，而BWC中仅有机号、预飞、推出三个字段的覆盖率超80%；
- 比较BWC与TBD各字段的首次更新时间，**TBD有更早更新的数据渠道**；
- 比较预计到达时间与实际到达时间的误差，BWC和TBD在误差分布上无明显差异，但TBD在数据总量上占绝对优势；
- BWC的实际起落时间值**60%以上与官方来源渠道不一致**；（官方指如航司、机场）

THANKS