

## Q&A RAS PSC 2015. 2 issue, Version of 13 July 2015

*Question 13): What is the difference between "LINE\_SEG\_NBR" and "TRACK\_SDTK\_NBR"? Is our understanding as shown below correct?*

*LINE\_SEG\_NBR: Defines from place A to place B*

- *TRACK\_SDTK\_NBR: Defines the track among the multiple tracks*

The interpretation is correct

*Question 14 ) In 'Training dataset', what do two rows in the sheet with same defect number mean? For example for 'Line segment' 3, 'Track sdtk nbr' 2, at 'Milepost' 287.24612 and 342.5195 the defect ID is same 1197. Tests were conducted on 14 September 2009 and 14 April 2008 respectively. The types of defects are 'Dip' and 'Surface' respectively.*

*Question 14) When does the same defect number get allocated to multiple instances in the data sheet?*

DEF\_NBR is not the unique key to identify a defect, you need to consider DEF\_NBR, EVT\_DT, GEO\_CAR\_NM.

*Question 15) What are the units for the column 'MILE POST'? (Miles/km)?*

Miles.

## Q&A RAS PSC 2015. Version of 8 July 2015

*Question 1): We found a puzzle about TONNAGE\_SAMPLE\_DATA in the RAS PSC - Training dataset file. The TONNAGE\_SAMPLE\_DATA only exists when LINE\_SEG\_NBR=4. However, we need all the TONNAGE\_SAMPLE\_DATA to obtain an algorithm. Is the data just like this or the other TONNAGE\_SAMPLE\_DATA hasn't been given?*

*Question 1) The TONAGE\_SAMPLE\_DATA has no information for LINE\_SEG\_NBR 1, 2, and 3. Did we miss anything?*

*Question 1) Why is tonnage data only provided for line 4?*

This was a data masking issue about the TONNAGE\_SAMPLE\_DATA. We updated the excel file with the correct data. (July 7)

*Question 2 ) The zip file in <https://www.informs.org/Community/RAS/Problem-Solving-Competition/2015-RAS-Problem-Solving-Competition> is the only set of data that will be provided for the competence?*

Yes. All data is available via the excel files.

*Question 3) Is there information of when a corrective action was performed in any point of the track?*

The text states that Repairs for yellow tag defects are not always recorded in the database. In general, thus, a red tag will trigger a corrective action, a yellow tag might, or might not, and this latter information is unavailable to the competition.

*Question 4) What is the maximum separation between measures for which it can be assumed that it is the same defect in the track?*

*Question 4) To track the degradation of a given defect, do successive runs need to mark it at the exact same spot?*

The text now reports on geographical repetition:

A defect found in the same location as a defect from the previous test may be flagged as a repeat defect. Defects must be from the same defect type and found within 100 feet on either side of a previous defect to be considered a repeat defect.

For temporal repetition, any two defects happening at the same location ( in the sense of the first half of this answer) , however far they are separated in time, they will be considered as the same.

We normally consider two inspection runs beyond 7 days' time window as consecutive inspections runs, if two inspection runs are within 7 days, they are considered in the same inspection run.

Even if a corrective action has been carried out, the yellow tags might not be 100% repaired. There are no maintenance data (correction data) available, since the corrective action for yellow tags might or might not recorded.

*Question 5) I have some ambiguities in the RAS problem competition data. In the overview you give an example of a data samples:*

*Yellow tag @milepost 12.000, 1 Feb*

*Yellow tag @milepost 12.000, 1 Apr*

*Yellow tag @milepost 12.000, 1 Sep*

*Red tag @Milepost 12.000, 18 Oct*

*The same milepost value is repeated in four samples but in the training dataset the inspection is done one time only for many milepost values. My question is: Can I consider the milepost "7,03336" the same as the milepost "7,03509" and if the differentiation of these values is related to columns milepost\_start and milepost\_end ?*

We say that :*"A defect found in the same location as a defect from the previous test may be flagged as a repeat defect. Defects must be from the same defect type and found within 100 feet on either side of a previous defect to be considered a repeat defect."*

*Question 6) Is Yellow/Red tag determined by defect amplitude alone or by amplitude and length together? Besides, what is the threshold for determining whether a certain type of track geometry defect is Yellow or Red? Our perception is that each railroad may have its own engineering standards to classify track defects (yellow versus red). We wonder if this information is available to us.*

*Question 6). What are the thresholds used in the training data for RED and YEL tag?*

How to determine precisely a yellow tag and red tag is sketched in the document. This information is actually not necessarily relevant, as the determination of yellow or red tag is

performed already in the data. The standards used might be different than the standard of a (another) particular railway company. Precisely, the following are the Yellow/Red limits to determine defect priority. However, these threshold are subject to change with time.

As of 09/01/2013	<b>Geometry Car Red and Yellow Tag Limits</b>				
	<b>Class 1</b>	<b>Class 2</b>	<b>Class 3</b>	<b>Class 4</b>	<b>Class 5</b>
<b>Max Speed</b>	<b>P15/F10</b>	<b>P30/F25</b>	<b>P60/F40</b>	<b>P80/F60</b>	<b>P90/F80</b>

	<b>Surface 62 (Left and Right)</b>				
<b>Red</b>	<b>3"</b>	<b>2 3/4"</b>	<b>2 1/4"</b>	<b>2"</b>	<b>1 1/4"</b>
<b>Yellow</b>	2 3/4"	2 1/2"	2"	1 3/4"	1"
	<b>Dip31</b>				
<b>Red</b>	<b>3"</b>	<b>2 3/4"</b>	<b>2 1/4"</b>	<b>1 3/4"</b>	<b>1 1/2"</b>
<b>Yellow</b>	2 1/2"	2 1/4"	1 3/4"	1 1/2"	1 1/4"

	<b>Crosslevel</b>				
<b>Red</b>	<b>3"</b>	<b>2"</b>	<b>1 3/4"</b>	<b>1 1/4"</b>	<b>1"</b>
<b>Yellow</b>	2 3/4"	1 3/4"	1 5/8"	1 1/8"	7/8"

*Question 7) There are four tracks in the problem. Are those tracks identical or they have different characteristics (e.g., rail age, engineering design)?*

The rail age and engineering design information are not available, but they are supposed to be different at least in location.

*Question 8) For the dip and cross level defects, we wonder why the same location can have two different defects measurement on the same date. For example, on row 419 & 420 of the training dataset (both from milepost 9 of line segment 2 on Apr 4, 2013) show two cross level defects at the same location. Similarly, row 2738 & 2739 (both from milepost 5 of line segment 2 on Aug 18, 2013) show two different dip defects at the same location (please see the table below).*

Row	LINE	MIL	TRAC	TES	DEF	GEO_	DEF	DE	DEF_	TS	CL	TES	TES	DFC
	_SEG	EPO	K_SD	T_D	_NB	CAR_	_PR	F_	AMP	C_	A	T_F	T_P	T_T
	_NB	ST	TK_N	T	R	NME	TY	LG	LTD	CD	SS	SPD	SPD	YPE
	R		BR					TH						
419	2	9	2	04A pr2 013	352 6	GEO5 04	RED	14	-1.27	T	4	55	60	XLE VEL

42 0	2	9	2	04A pr2 013	352 7	GEO5 04	YEL	21	-1.24	T	4	55	60	XLE VEL
27 38	2	5	0	18A ug2 013	594 59	GEO1 07	YEL	10	1.69	T	4	50	55	DIP
27 39	2	5	0	18A ug2 013	594 60	GEO1 07	YEL	4	1.82	T	3	40	40	DIP

*Question 8). Data shows several instances where two SURF defects are shown at the same spot on the same day. Are these for Left and Right rail independently?*

The track geometry car might run across the same location multiple times at the same day. Two or more inspection runs within 7 days (including the same day) are considered the same inspection run.

Moreover, the defect might come from the left and right rail or come from the same rail.

*Question 9) From the problem statement, we understand that the surface defect can be either on the left rail or the right rail. In the training dataset, is there any information tell us whether a defect is on the left versus right rail?*

*Question 9). Data shows several instances where two SURF defects are shown at the same spot on the same day. Are these for Left and Right rail independently?*

No. The track ( both rails) will be deemed as yellow tag or red tag, if any or both of the two rails are not according the standard.

*Question 10) How is defect length determined? How to report the milepost within defect length (e.g., the milepost is the middle point of the defect length, or it is the starting or ending point of defect length?)*

The milepost is the location where the defect is measured the worst

*Question 11) Again in the TONAGE\_SAMPLE\_DATA, some rows contain total gross tons but do not have information regarding total number of cars traveling east/west, or zero total trains traveling east/west. Do we have to infer the number of trains and cars?*

Participants are free to either infer the number of trains and cars or not.

*Question 12) Many mileposts have only one inspection record within the study period. Therefore, we consider aggregating some small segments in the training dataset into a larger track lot. Is there any specified aggregation length we need to use (e.g., 0.01 mile or 0.1 mile?)*

No fixed specified aggregation length is suggested.

*Question 13) What is the 'length' of a defect? Is it measured +/-0.5Length, or point to point + Length?*

We are not provided the detailed information how to get the “length”. The operator uses the length which is beyond the yel/red threshold, and based on the consecutive yel/red tags, they might chain or group some tags together to have a combined defect length.

*Question 14). Why are SURF defects all negative and DIP all positive?*

Theoretically, both SURF and DIP can be either negative or positive. The sampled dataset is consistent with the original dataset.

*Question 15) There seem to be inconsistencies in inspection mileposts data. Ideally, there should not be any other track over span of a single track (TRACK\_SDTK\_NBR = 0). But, the data set have those entries. For example line1, track segment 0, 1 and 2 spans over 0-226.09, 199.28-203.2, 199.28-203.2 respectively*

It is not inconsistencies, it is due to the parallel track segments.