**.3GPP TSG-RAN WG2 Meeting #127bis** **R2-240xxxx**

**Hefei, China, Oct 14th~ Oct 18th, 2024**

Agenda Item: 8.3.2

Source: Mediatek Inc., OPPO

Title: [POST128][021][AI Mob] Templates for simulations (Mediatek/OPPO)

Document for: Discussion, Decision

# Introduction

This report provides a summary for the following post-meeting email discussion:

* [POST128][021][AI Mob] Templates for simulations (Mediatek/Oppo)

Intended outcome:

1. agree to updated and new templates. Companies should update their simulation results in the new template (Mediatek)

2. agree principles on how to capture observations from simulations (Oppo)

Deadline: Jan. 15th

The deadline of the email discussion is Jan. 15th. Please provide your comment by Jan. 15th, 8:00 UTC to allow us sufficient time to revise the table and provide the summary.

Companies providing input to this email discussion are requested to leave contact information below.

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **Email Address** |
| NTT DOCOMO | Xin Wang | wangx@docomolabs-beijing.com.cn |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Simulation Template

## Revisions on the templates for RRM Prediction

### Revisions for Sub-use case 2

According to the agreement ‘Company can report which filtering options is being used for the input L3 RSRP of sub-use case 2: option 1, option 2, option 3.’, the filtering options for sub-use case 2 need to be reported. One simple method is to define the format of the content for the sub-use case column as 1/2\_FilteringOption1/2\_FilteringOption2/2\_FilteringOption3/3.

**Q1: Do you agree to define the format of the content for the sub-use case column as**  **1/****2\_FilteringOption1/2\_FilteringOption2/2\_FilteringOption3/3, where 2\_FilteringOption1, 2\_FilteringOption2, and 2\_FilteringOption3 corresponds to filtering options 1, 2, and 3 for sub-use case 2?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Generalization for RRM Predication

In RAN2#128 meeting, we made following agreements for generalization performance evaluation:

|  |
| --- |
| **Agreements on generalization**   1. Reuse the evaluation methodology in TR38.843 for generalization study, i.e., the generalization performance is evaluated with the following cases,  * *Baseline:* The AI/ML model is trained using the dataset with Configuration #B and tested using the dataset with Configuration #B. * *Generalization Case #1 (GC#1):* The AI/ML model is trained using the dataset with Configuration #A but tested using the dataset with Configuration #B. * *Generalization Case #2 (GC#2):* The AI/ML model is trained using mixed datasets with both configurations and tested using the dataset with Configuration #B.   2 Companies can choose which case they compare with and should report it with simulation results.  3 Generalization issues on RRM measurement prediction are prioritized.  4 Start the study with generalization issue with RRM measurement prediction in temporal domain. Companies can chose to study frequency domain prediction cases and report what they have simulated.  5 Study generalization over UE speeds  6 The simulation assumption of FR1 temporal domain case B is reused for generalization study with 3 UE speeds i.e. 30Km/h, 60Km/h and 90Km/h. FFS on combinations  7 The simulation assumption of FR2 temporal domain case A is reused for generalization study with 3 UE speeds i.e. 60Km/h, 90Km/h and 120Km/h. FFS on combinations |

For evaluating generalization performance, it is recommended to use a separate spreadsheet distinct from the one used for non-generalization performance. Separate spreadsheets for generalization performance evaluation are introduced for Scenario 2 and Scenario 4, as RAN2 has agreed to start the study in the temporal domain.

**Q2: Do you agree to introduce a separate spreadsheet for generalization performance evaluation in Scenario 2 and Scenario 4?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes | The spreadsheet for Scenario 3 should also be introduced since some companies will also study it. The generalization between 2GHz->4GHz and 4GHz->2GHz can be included at least.  We have noticed that there is a template for it and it should be kept to collect the results. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

We have agreed to study model generalization across different cell configurations (e.g., ISD, gNB height, power, beam pattern, etc.). It remains to be decided which parameters we will prioritize. We will have another email discussion [Post128][018][AI Mob] Generalization (Apple) to discuss the parameters for different cell configurations. The outcome of this discussion will determine whether the generalization performance should be evaluated for each individual parameter or combinations of those parameters.

According to the agreed principles of the evaluation methodology for the generalization study, it is recommended to add a category for generalization-related information. This category should include information and columns for generalization cases (GC1/GC2), Configuration A, and Configuration B. Configuration A refers to the dataset configuration used to train the AI/ML model, while Configuration B refers to the dataset configuration used for inference.

For example, in the generalization study for FR1 with UE speed:

* For GC#1, Configuration A is 30 km/h, while Configuration B is 60 km/h or 90 km/h.
* For GC#2, Configuration A is a combination of 30 km/h, 60 km/h, and 90 km/h, while Configuration B is 30 km/h.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Training @Dataset: 30km/h* | *Training @Dataset: 60km/h* | *Training @Dataset: 90km/h* | *Inference @30km/h* | *Inference @60km/h* | *Inference @90km/h* |
| *Baseline* | *Yes* |  |  | *Yes* |  |  |
| *GC#1* | *Yes* |  |  |  | *Yes* | *Yes* |
| *GC#2* | *Yes* | *Yes* | *Yes* | *Yes* |  |  |

**Table 1**

Therefore, Configuration A and B refer to either a single parameter or a combination of parameters, each having no more than 2 values per parameter, as determined by the outcome of the email discussion [Post128][018][AI Mob] Generalization.

We will have separate sheets for different parameters or combinations of parameters for the generalization performance evaluation. For example, one sheet for UE speed, one sheet for beam pattern (if agreed), one sheet for ISD (if agreed), etc.

**Q3:** **Do you agree to add a category for generalization-related information, which includes columns for generalization cases (Baseline/GC1/GC2), Configuration A, and Configuration B? Configuration A refers to the dataset configuration used to train the AI/ML model, while Configuration B refers to the dataset configuration used for inference. Both configurations can consist of a single parameter or a combination of different parameters.?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Q4: Do you agree to have separate sheets for different parameters or combinations of parameters for the generalization performance evaluation? For example, one sheet for UE speed, one sheet for beam pattern (if agreed), one sheet for ISD (if agreed), etc.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Generally Yes | We are general fine with the principle and creating one sheet for UE speed. For the remaining issues, it depends on the conclusion of the post email discussion [Post128][018] (whether we treat these factors separately or jointly). |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## System-Level Variable Setting and Metrics for Measurement Event Prediction

### Spreadsheets for Measurement Event Prediction

For measurement event prediction, we agreed in RAN2#127bis meeting that *measurement event prediction simulations will at least focus on intra-frequency FR2, case A, and second study goal (i.e. HO KPI improvement). Companies can bring simulation results for intra-frequency measurement reduction for FR1 and report what they are doing. Focus on temporal case B. Companies will prioritize simulations on indirect method. Companies can bring simulations on direct method and should report what method is being used.*

For system level simulation, we agreed in RAN2#127bis meeting that *the system level performance (e.g. HO performance) evaluation is optional (i.e. companies can bring results if they chose). System level performance for measurement event prediction can be prioritized by companies if they chose to do it.*

It is recommended that we have two sheets corresponding to indirect and direct measurement event prediction for both Case A and Case B. To clarify the impact of AI model performance (i.e. intermediate KPIs) on system performance, the system level metrics the intermediate KPIs should be placed on the same sheet. For each spreadsheet, the intermediate KPIs agreed upon for measurement event prediction, including F1 score, Precision, and Recall, will be added. For indirect prediction, the KPIs for RSRP difference will also be included and optionally reported and maxETD will be added as a variable setting.

**Q5: Do you agree to have separate sheets corresponding to indirect and direct measurement event prediction for both Case A and Case B, with the intermediate KPIs (including F1 score, Precision, and Recall) and system level metrics (including HO failure rate and total number of HO attempts per UE per second) added?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Q6: Do you agree to keep the intermediate KPIs of RSRP difference for indirect prediction, which is optionally reported by companies?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

### Measurement Event Prediction-Case B

For handover modelling for measurement event prediction for temporal domain case B, we agreed that *UE reports when A3 event is satisfied with actual measurements and predicted results. And handover command will be received after handover preparation.*

*Agreed values for Case B. NOTE1 indirect prediction only*

|  |  |  |
| --- | --- | --- |
| *Parameters* | *baseline value* | *Note* |
| *A3 event offset (db)* | *2* | *Open for 3db* |
| *TTT (ms)* | *320* | *Open for one shorter value* |
| *UE speed (km/h)* | *30* | *Open for 60 and 90km/h* |
| *OW length (ms,note1)* | *N/A* | *Up to implementation* |
| *PW length (ms,note1)* | *200 (non-sliding)*  *40ms (sliding)* | *Open for more values* |
| *Max ETD (ms, note1)* | *80* | *Open for more values* |
| *MRRT* | *50%* | *Open for more values* |

**Table 2**

Based on the agreement, the parameters of A3 event offset and TTT can be considered as variable settings for system level simulation.

**Q7: Do you agree to add the parameters of A3 event offset and TTT as variable settings for indirect measurement event prediction for Case B?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

### Measurement Event Prediction-Case A

For handover modelling for measurement event prediction for temporal domain case A, we agreed that company focus on option 2 or option 3.

* *Option 2: network transmit handover command purely based on actual measurement event regardless whether an actual measurement result(@t2) is earlier or later than predicted measurement event((@t1))*
* *Option 3: For AI mobility, HO preparation starts when an event is predicted to happen (i.e., t0), and HO command is sent when A3 entering conditions are met based on actual/real measurement and an event is predicted to be met for the duration of TTT.*

*Agreed values for Case A. NOTE1 indirect prediction only*

|  |  |  |
| --- | --- | --- |
| *Parameters* | *baseline value* | *Note* |
| *A3 event offset (db)* | *2* | *Open for 3db* |
| *TTT (ms)* | *320* | *Open for one shorter value* |
| *UE speed (km/h)* | *90* | *Open for 60 and 120km/h* |
| *OW length (ms, note1)* | *N/A* | *Up to implementation* |
| *PW length (ms, note1)* | *400* | *Open for more values* |
| *Max ETD (ms, note1)* | *80* | *Open for more values* |

**Table 3**

Same as Case A that A3 event offset and TTT are added as the variable settings. In addition, the handover model option2/option 3 is also added.

**Q8: Do you agree to add the parameters of A3 event offset and TTT as variable settings in addition to handover modelling (option 2/option3) for indirect measurement event prediction for Case A?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Comments for the spreadsheet examples

For February meeting, the rapporteur will provides following 9 templates to collect simulation results.

* Example1.1\_Scenario 2: RRM Measurement Prediction Evaluation results for intra-frequency temporal domain case B without generalization;
* Example1.2\_Scenario 2: RRM Measurement Prediction Evaluation results for intra-frequency temporal domain case B with generalization;
* Example 2.1\_Scenario 4: RRM Measurement Prediction Evaluation results for intra-frequency temporal domain case A without generalization;
* Example 2.2\_Scenario 4: RRM Measurement Prediction Evaluation results for intra-frequency temporal domain case A with generalization;
* Example 3.1\_Scenario 3: RRM Measurement Prediction Evaluation results for inter-frequency (frequency domain) without generalization;
* Example 3.2\_Scenario 3: RRM Measurement Prediction Evaluation results for inter-frequency (frequency domain) with generalization;
* Example 4\_Scenario 6: RRM Measurement Prediction Evaluation results for intra-frequency spatial domain without generalization.
* Example 5\_ Case B: Measurement Event Prediction Evaluation results;
* Example 6\_Case A: Measurement Event Prediction Evaluation results.

**Q9: Do you agree to have the above templates to collect the simulation results for February Meeting?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Q10: Please provide your comments on each template example in the following tables.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Company** | **Example1.1** | **Example1.2** | **Example 2.1** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Company** | **Example 2.2** | **Example 3.1** | **Example 3.2** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Company** | **Example 4** | **Exampel 5** | **Example 6** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Files Naming

Since we have new tables for measurement event prediction and generalization, and some companies have indicated that the original naming cannot distinguish different scenarios from the document's naming, we propose the following new naming rules:

|  |  |  |
| --- | --- | --- |
| Template file name | Submit file format | Submit file example |
| Example 1.1\_Scenario 2 | RRM\_Scen2\_[Meeting nubmer]\_[Company]\_[Tdoc number]\_[ver].xlsx | RRM\_Scen2\_127bis\_Mediatek\_2409869\_v1.xlsx |
| Example 1.2\_Scenario 2 | RRM\_Scen2\_Gen\_[Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx | RRM\_Scen2\_Gen\_129\_Mediatek\_2xxxxxx\_v1.xlsx |
| Example 2.1\_Scenario 4 | RRM\_Scen4\_[Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx | RRM\_Scen4\_127bis\_Mediatek\_2409869\_v1.xlsx |
| Example 2.2\_Scenario 4 | RRM\_Scen4\_Gen\_[Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx | RRM\_Scen4\_Gen\_129\_Mediatek\_2xxxxxx\_v1.xlsx |
| Example 3.1\_Scenario 3 | RRM\_Scen3\_[Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx | RRM\_Scen2\_127bis\_Mediatek\_2409869\_v1.xlsx |
| Example 3.2\_Scenario 3 | RRM\_Scen3\_Gen\_[Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx | RRM\_Scen3\_Gen\_129\_Mediatek\_2xxxxxx\_v1.xlsx |
| Example 4\_Scenario 6 | RRM\_Scen6\_[Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx | RRM\_Scen6\_127bis\_Mediatek\_2409869\_v1.xlsx |
| Example 5\_ Case B\_Indirect | ME\_CaseB\_ [Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx  (Note: ME refer to measurement event) | ME\_CaseB\_ 129\_Mediatek\_2xxxxxx\_v1.xlsx |
| Example 6\_Case A\_Indirect | ME\_CaseA\_ [Meeting nubmer]\_[Company]\_[Tdoc num] \_[ver].xlsx | ME\_CaseA\_ 129\_Mediatek\_2xxxxxx\_v1.xlsx |

**Table 4**

**Q11: Do you agree with the naming rules of the submitting Excel tables?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| NTT DOCOMO | Yes |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

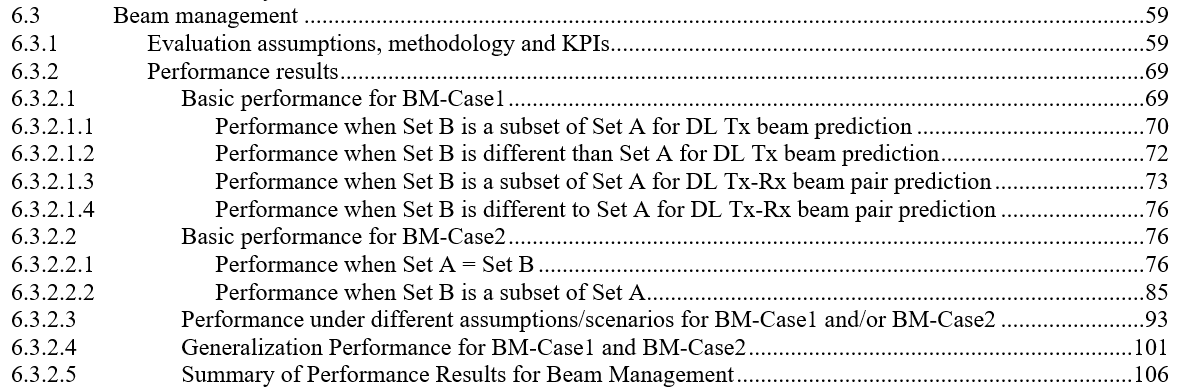
## Other Comment

**Q12: Do you have any other suggestions on the templates and process of collecting the simulation results?**

|  |  |
| --- | --- |
| **Company** | **Comment** |
| NTT DOCOMO | The method of drawing quantitative observations from the results submitted by companies should be further discussed (maybe during the next RAN2 meeting). The issues to be address includes which type of results will be used to draw the conclusion (abolute performance or relative gain), and which statistics should be calculated from the collected results, e.g., mean/medium values, quantile values, standard deviations, etc. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Principle to capture simulation result

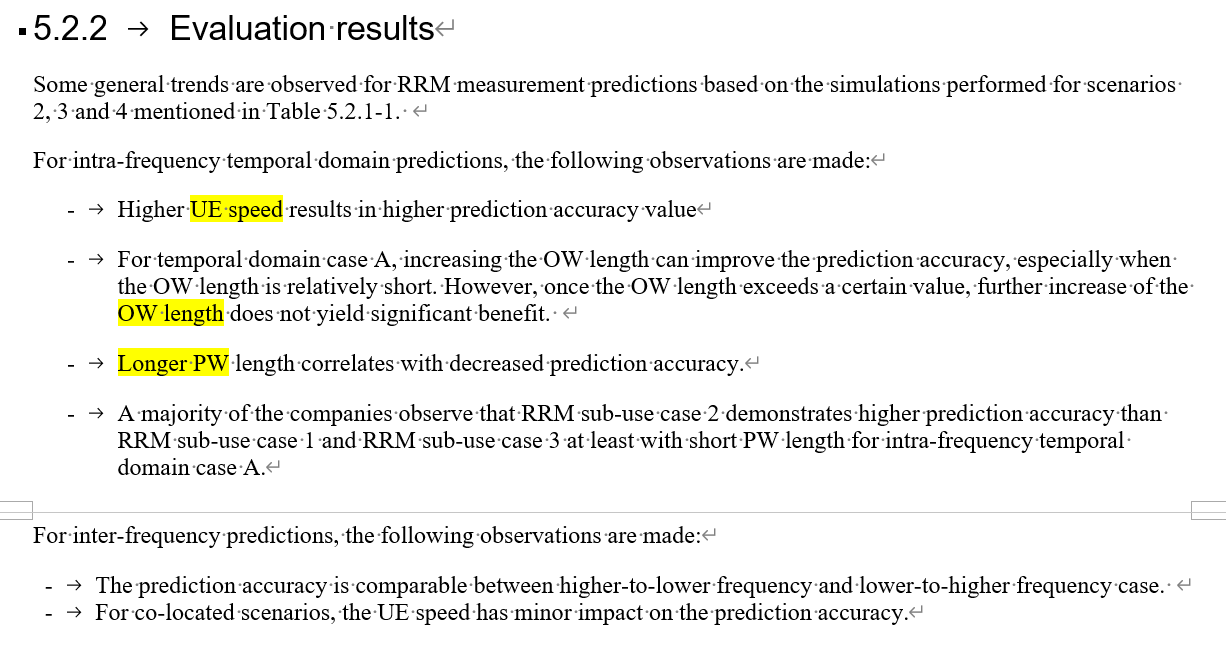
In TR 38.843 [1], companies’ simulation results under different scenarios are captured in different sub-sections. The ways to capture performance results for beam management are illustrated in the following structure:



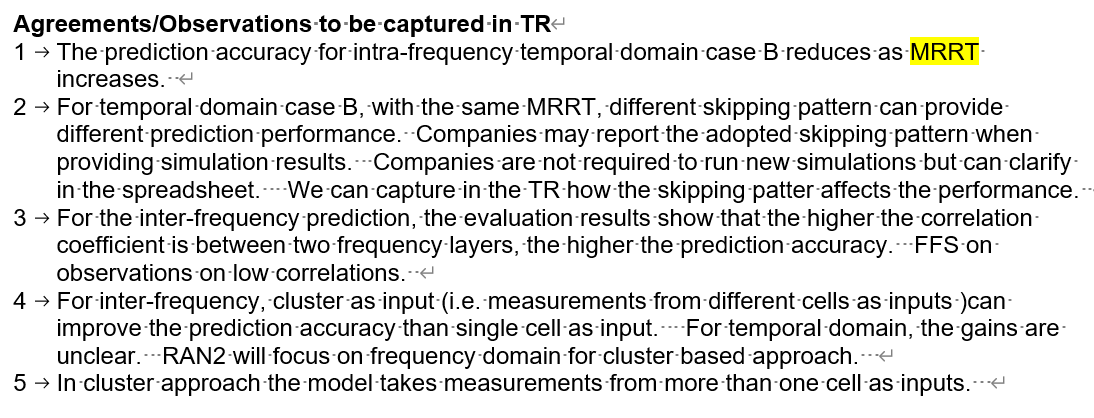
**Figure 1. Illustration of the hierarchy of beam management section in TR 38.843**

We can see from Figure 1 that the basic principle is to give key cases a separate sub-section to record the detailed results and leave the rest to the “performance under different assumptions/scenarios…” sub-section.

In RRM, the current simulation results focus on high- and medium- priority scenarios, i.e., scenarios 2, 3, 4 and 6. If we follow the same principle as TR 38.843, we can give those higher-priority scenarios separate sub-sections to capture their extensive simulation results. For each high-priority scenario, we have simulated the performance under different sub-use cases. The input and output of different sub-use cases vary and may have different specification impacts. Therefore, it is better to place them into parallel sections secondary to the basic performance of scenarios. For the four higher-priority scenarios, the factors that impact the performance results vary and should be treated case by case. Here is what is captured in the TR:



During this RAN2#128 meeting, further general observations are agreed as following:



Based on those agreements, the key parameters and corresponding scenario is summarized in following table:

|  |  |
| --- | --- |
| Scenario | Key parameters |
| Scenario 2 | MRRT, UE speed |
| Scenario 3 | channel correlation coefficient, cluster |
| Scenario 4 | OW, PW, UE speed |
| Scenario 6 | MRRS, UE speed |

**Table 5**

For scenario 3 (frequency domain), it is likely that RAN2 focus on 2GHz and 4GHz i.e. there is no more pair of frequencies to be compared w.r.t. channel corelation coefficient. While cluster approach could play a role. To differentiate between single cell and cluster approach, the simulation results can be further categorized into single cell approach and cluster approach.

For scenario 4, both OW and PW matters. But it sounds like PW is more important. It will be very difficult to list all the OW vs PW combinations. Instead, we can list few PW lengths.

For scenario 6, even MRRS and UE speed will impact the performance, it is not necessary to differentiate simply because the simulation result is not so much.

By taking all the above agreements and analysis into account, here is recommended structure to capture simulation observation from collected simulation results:

Basic performance for scenario 2

Sub use case 1/2/3

MRRT, UE speed

Basic performance for scenario 3

Sub use case 1/2/3

Single cell, cluster

Basic performance for scenario 4

Sub use case 1/2/3

PW, UE speed

Basic performance for scenario 6

Generalization performance

Summary of performance results

You can find an example in the Annex part.

Company may have different view on how to select the key parameters to reflect the performance in the last level as illustrated above. If you want to suggest some other parameters, please refer to parameters RAN2 has ever discussed and concluded for general observation as much as possible to make the observation meaningful and brief. In any case, all the simulation results will be captured in the spreadsheet and they will be part of the TR also.

**Q13: Do you agree the recommended structure to capture evaluation observations for collected simulation results?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/no | comments |
| NTT DOCOMO | Yes | We generally agree the principles and structures suggested by the FL. Like the existing use cases in TR 38.843, we can say xx sources observed the performance in a certain range as detailed observations for each case and subcase.  In addition, like we discussed in Q12, we also need a high-level and summarized performance gain to conclude the SI. How to summarize the results from companies can be further discussed, e.g., use the averaged or medium RSRP gap, or capture the quantiles (used in RAN1 AI/ML CSI prediction study) of the results from different companies, something like that. |
|  |  |  |

# Conclusion

# Annex- Proposed structure for RRM evaluation results

5.2.2 Evaluation results

5.2.2.1 Basic performance for scenario 2

5.2.2.1.1 Performance of sub case 1

Table 5.1 example for capturing performance results

|  |  |  |  |
| --- | --- | --- | --- |
| UE speed \ MRRT | 50% | 67% | … |
| 30km/h | * Average L3 RSRP cell difference   [0.1, 1] dB: 9 sources  [xx, xx] dB: xx sources   * … | * … |  |
| 60km/h | * … | * … |  |
| … |  |  |  |

5.2.2.1.2 Performance of sub case 2

5.2.2.1.3 Performance of sub case 3

5.2.2.2 Basic performance for scenario 3

5.2.2.2.1 Performance of sub case 1

Single cell approach:

* Average L3 RSRP cell difference

[0.1, 1] dB: 9 sources

[xx, xx] dB: xx sources

…

Cluster approach:

5.2.2.2.2 Performance of sub case 2

5.2.2.2.3 Performance of sub case 3

5.2.2.3 Basic performance for scenario 4

5.2.2.3.1 Performance of sub case 1

Table 5.2 example for capturing performance results

|  |  |  |  |
| --- | --- | --- | --- |
| UE speed \ PW | 200ms | 400ms | … |
| 60km/h | * Average L3 RSRP cell difference   [0.1, 1] dB: 9 sources  [xx, xx] dB: xx sources   * … | * … |  |
| 90km/h | * … | * … |  |
| … |  |  |  |

5.2.2.3.2 Performance of sub case 2

5.2.2.3.3 Performance of sub case 3

5.2.2.4 Basic performance for scenario 6

* + Average L3 RSRP cell difference

[0.1, 1] dB: 9 sources

[xx, xx] dB: xx sources

5.2.2.5 Generalization performance

5.2.2.6 Summary of performance results