|  |
| --- |
| 3GPP TR 38.744 V1.2.0 (2025-9) |
| Technical Report |
| 3rd Generation Partnership Project;Technical Specification Group Radio Access Network;Study on Artificial Intelligence (AI)/Machine Learning (ML) for Mobility in NR;(Release 19) |
|   |
|  |  |
|  |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented.This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification.Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. |

|  |
| --- |
|  |
| ***3GPP***Postal address3GPP support office address650 Route des Lucioles - Sophia AntipolisValbonne - FRANCETel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16Internethttps://www.3gpp.org |
| ***Copyright Notification***No part may be reproduced except as authorized by written permission.The copyright and the foregoing restriction extend to reproduction in all media.© 2024, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).All rights reserved.UMTS™ is a Trade Mark of ETSI registered for the benefit of its members3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersLTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersGSM® and the GSM logo are registered and owned by the GSM Association |

Contents

[Foreword 6](#_Toc207617049)

[1 Scope 8](#_Toc207617050)

[2 References 8](#_Toc207617051)

[3 Definitions of terms, symbols and abbreviations 8](#_Toc207617052)

[3.1 Terms 8](#_Toc207617053)

[3.2 Abbreviations 9](#_Toc207617054)

[4 AI/ML mobility use cases 9](#_Toc207617055)

[4.1 General 9](#_Toc207617056)

[4.2 RRM measurement prediction 9](#_Toc207617057)

[4.3 Measurement event prediction 10](#_Toc207617058)

[4.4 RLF prediction 11](#_Toc207617059)

[5 Evaluations 11](#_Toc207617060)

[5.1 Common evaluation methodology, metrics and assumptions 11](#_Toc207617061)

[5.2 RRM measurement prediction 15](#_Toc207617062)

[5.2.1 Evaluation methodology, metrics and assumptions 15](#_Toc207617063)

[5.2.1.1 RRM measurement prediction 15](#_Toc207617064)

[5.2.1.2 Generalization 17](#_Toc207617065)

[5.2.2 Evaluation results 18](#_Toc207617066)

[5.2.2.1 RRM measurement prediction 19](#_Toc207617067)

[5.2.2.1.1 Basic performance for FR1 intra-frequency temporal domain case B 19](#_Toc207617068)

[5.2.2.1.2 Basic performance for FR1 inter-frequency prediction 21](#_Toc207617069)

[5.2.2.1.3 Basic performance for FR2 intra-frequency temporal domain case A 21](#_Toc207617070)

[5.2.2.1.4 Summary of performance results for RRM measurement prediction 22](#_Toc207617071)

[5.2.2.2 Generalization 23](#_Toc207617072)

[5.2.2.2.1 Generalization performance for FR1 intra-frequency temporal domain case B 23](#_Toc207617073)

[5.2.2.2.2 Generalization performance for FR1 inter-frequency prediction 24](#_Toc207617074)

[5.2.2.2.3 Generalization performance for FR2 intra-frequency temporal domain case A 25](#_Toc207617075)

[5.2.2.2.4 Summary of performance results for generalization of RRM measurement prediction 26](#_Toc207617076)

[5.3 Measurement event prediction 27](#_Toc207617077)

[5.3.1 Evaluation methodology, metrics and assumptions 27](#_Toc207617078)

[5.3.2 Evaluation results 28](#_Toc207617079)

[5.3.2.1 Performance of measurement event prediction based on FR2 intra-frequency temporal domain case A 28](#_Toc207617080)

[5.3.2.2 Performance of measurement event prediction based on FR1 intra-frequency temporal domain case B 29](#_Toc207617081)

[5.3.2.3 Summary of performance results for measurement event prediction 29](#_Toc207617082)

[5.4 RLF prediction 30](#_Toc207617083)

[5.4.1 Evaluation methodology, metrics and assumptions 30](#_Toc207617084)

[5.5 System level simulation 31](#_Toc207617085)

[5.5.1 Evaluation methodology, metrics and assumptions 31](#_Toc207617086)

[5.5.2 Evaluation results 32](#_Toc207617087)

[5.5.2.1 SLS Performance of measurement event prediction based on FR2 intra-frequency temporal domain case A 32](#_Toc207617088)

[5.5.2.2 SLS Performance of measurement event prediction based on FR1 intra-frequency temporal domain case B 33](#_Toc207617089)

[5.5.2.3 Summary of SLS Performance 34](#_Toc207617090)

[6 Potential specification impact 34](#_Toc207617091)

[6.1 LCM, protocol and procedure aspects 34](#_Toc207617092)

[6.1.1 Overview 34](#_Toc207617093)

[6.1.2 RRM measurement prediction 35](#_Toc207617094)

[6.1.2.1 UE-sided model 35](#_Toc207617095)

[6.1.2.1.1 Applicability reporting 35](#_Toc207617096)

[6.1.2.1.2 Inference configuration and report 35](#_Toc207617097)

[6.1.2.1.3 Monitoring and management 36](#_Toc207617098)

[6.1.2.1.4 Data collection for offline training 36](#_Toc207617099)

[6.1.2.2 Network-sided model 36](#_Toc207617100)

[6.1.2.2.1 Inference input reporting 36](#_Toc207617101)

[6.1.2.2.2 Monitoring and management 37](#_Toc207617102)

[6.1.2.2.3 Data collection for offline training 37](#_Toc207617103)

[6.1.3 Measurement event prediction 37](#_Toc207617104)

[6.2 Interoperability, testability and RRM requirements 38](#_Toc207617105)

[6.2.1 RRM requirements for measurement prediction 38](#_Toc207617106)

[6.2.1.1 General 38](#_Toc207617107)

[6.2.1.2 Potential RRM requirements 38](#_Toc207617108)

[6.2.2 RRM requirements for measurement event prediction 39](#_Toc207617109)

[6.2.2.1 General 39](#_Toc207617110)

[6.2.2.2 Potential RRM requirements 39](#_Toc207617111)

[6.2.3 Testability for RRM measurement prediction 40](#_Toc207617112)

[6.2.3.1 Testing goal 40](#_Toc207617113)

[6.2.3.2 Prediction consistency in time domain 40](#_Toc207617114)

[6.2.3.3 Testing setup 40](#_Toc207617115)

[6.2.4 Interoperability 40](#_Toc207617116)

[6.2.5 Generalization 40](#_Toc207617117)

[7 Conclusion 40](#_Toc207617118)

[Annex <A> Change history: 41](#_Toc207617119)

# Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# 1 Scope

The study on physical layer use cases captured in [5] shows potentials for AI/ML algorithm in cellular communication system. Further study on use case of AI mobility is hence conducted in this study.

This study explores RRM measurement and measurement event prediction mainly in temporal domain (FR1) and frequency domain (FR2) to understand the feasibility and performance of AI/ML algorithm for measurement reduction or handover performance improvement based on simulation evaluation. Analysis on specification impact is carried out for both UE sided model and network sided model for said scenarios.

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 38.331: "NR; Radio Resource Control (RRC); Protocol specification".

[3] 3GPP TS 38.133: "NR; Requirements for support of radio resource management".

[4] 3GPP TR 38.901: "Study on channel model for frequencies from 0.5 to 100 GHz"

[5] 3GPP TR 38.843: “Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface”

[6] 3GPP TS 38.300: “NR and NG-RAN Overall description; Stage-2”

[7] 3GPP TR 36.839: “Mobility enhancements in heterogeneous networks”

# 3 Definitions of terms, symbols and abbreviations

## 3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

## 3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

<ABBREVIATION> <Expansion>

ETD Time Distance of measurement Events

GC Generalization Case

HOF Handover Failure

MRRS Measurement Reduction Rate in Spatial domain

MRRT Measurement Reduction Rate in Temporal domain

OW Observation Window

PCI Physical Cell Identity

PW Prediction Window

RLF Radio Link Failure

SLS System Level Simulation

UAI UE Assistant Information

# 4 AI/ML mobility use cases

## 4.1 General

The use cases in this study focus on RRC\_CONNECTED mode and cover RRM measurement prediction, measurement event prediction and RLF/HOF prediction for PCell and/or SCell change procedure in standalone NR scenario. The study of the use cases is driven mainly by two study goals. The 1st study goal is to reduce measurement efforts in temporal, spatial or frequency domain by using predicted measurements. The 2nd study goal is to improve the handover performance (e.g., Ping-pong HO, HOF/RLF, short time of stay, Handover interruption).

## 4.2 RRM measurement prediction

3 sub-use cases are considered for cell-level RRM measurement prediction:

- RRM sub-use case 1: L1 beam-level measurement result(s) is predicted based on actual L1 beam-level measurement result(s) and then L3 cell-level measurement result is generated;

- RRM sub-use case 2: L3 Cell-level measurement result(s) is predicted based on actual L3 cell-level measurement result(s);

- RRM sub-use case 3: L3 Cell-level measurement result(s) is predicted based on actual L1 beam-level measurement result(s).

3 sub-use cases are considered for beam-level RRM measurement prediction:

- Sub-use case 4: L1 filtered beam-level measurement result(s) is predicted based on actual L1 beam-level measurement result(s) and then L3 beam-level measurement result is generated;

- Sub-use case 5: L3 beam-level measurement result(s) is predicted based on actual L3 beam-level measurement result(s);

- Sub-use case 6: L3 beam-level measurement result(s) is predicted based on actual L1 beam-level measurement result(s).

For intra-frequency temporal domain case B (defined in section 5.2.1.1), there are 3 filtering options as for the input of RRM sub-use case 2 if immediate last measurement result(s) is skipped:

- Filtering option 1: L3 filtering is based on its L1 filtered result and the immediate last skipped measurement result;

- Filtering option 2: L3 filtering is based on its L1 filtered result, i.e. no L3 filtering;

- Filtering option 3: L3 filtering is based on the L1 filtered result and last actual measurement result, i.e. the skipped result(s) in between is ignored.

The skipped result refers to L3 RSRP measurement result predicted previously by the RRM measurement prediction model.

NOTE1: Actual measurement result refers to historical measurement result obtained using the legacy measurement framework

## 4.3 Measurement event prediction

There are two methods to predict measurement event, namely indirect and direct measurement event prediction as illustrated in Figure 4.3-1 and Figure 4.3-2 respectively.



Figure 4.3-1: Indirect measurement event prediction

In indirect measurement event prediction for intra-frequency temporal domain case A, temporal domain case B or spatial domain, measurement result(s) is predicted by a RRM measurement prediction model at first. Afterwards, predicted and optionally actual historical measurement result(s) of the same cell(s) are used to derive whether a measurement event at one future time instance occurs, without further involvement of an AI/ML model.

In indirect measurement event prediction for frequency domain, measurement result(s) is predicted by a RRM measurement prediction model for frequency domain at first. Afterwards, predicted and optionally actual historical measurement result(s) of serving cell are used to derive whether a measurement event at one time instance occurs, without further involvement of an AI/ML model.



Figure 4.3-2: Direct measurement event prediction

As illustrated in Figure 4.3-2, the input of the model with direct prediction is the same as indirect prediction as illustrated in Figure 4.3-1 and additional input is also allowed for both. Measurement event is predicted directly by an AI/ML model, i.e. the output of the model is the likelihood of an event occurrence.

For measurement event prediction based on intra-frequency temporal domain case B, the 3 filtering options captured in section 4.2 also apply for the input of RRM sub-use case 2. For indirect prediction, the skipped result refers to L3 RSRP measurement result predicted previously by the RRM measurement prediction model. For direct prediction, the skipped result refers to skipped L1 filtered measurement result and filtering option 1 is not applicable.

## 4.4 RLF prediction

The study focuses on RLF detected upon T310 expiry in PCell [2].

RLF can be predicted indirectly or directly based on actual measurement result(s) e.g. L1-SINR of PCell as illustrated in Figure 4.4-1 and Figure 4.4-2 respectively. In indirect RLF prediction, the future L1 SINR results are predicted based on actual historical L1 SINR results of the serving cell. Afterwards, RLF event at future time instance is determined based on predicted and optionally actual L1-SINR results within T310 duration, without further involvement of an AI/ML model. As baseline L1-SINR refers to raw L1-SINR without L1 filtering.



Figure 4.4-1: Indirect RLF prediction

In direct RLF prediction the likelihood of an RLF is predicted based on actual measurements (e.g. L1-SINR of PCell) directly.



Figure 4.4-2: Direct RLF prediction

# 5 Evaluations

## 5.1 Common evaluation methodology, metrics and assumptions

Synthesized datasets based on channel model and deployment [4] are used for evaluation. Field data can be used optionally. In principle once a set of simulation parameters and assumptions are settled, it should also be used for the baseline case (i.e. without AI/ML model), model training (e.g. data set generation), model validation, model test and inference operation [5] etc. Between training and test data set, different random seeds are used at least for channel modelling and UE trajectory. No traffic model is simulated in this study.

Both sliding L1/L3 filtering and non-sliding L1/L3 filtering options can be used for evaluation.



Figure 5.1-1: Sliding L1/L3 filtering



Figure 5.1-2: Non-sliding L1/L3 filtering

In sliding L1/L3 filtering, filtered L1 or L3 measurement result are generated every sample period. In non-sliding L1/L3 filtering, filtered L1 or L3 measurement result are generated every measurement period.

In both L1/L3 filtering options, the filtered L1 measurement result is obtained based on the raw L1 measurement results corresponding to reference point A in Figure 9.2.4-1 in [6] within one measurement period. The filtered L3 measurement result is obtained as specified in section 5.5.3.2 of [2].

In cluster approach, measurement results from more than one cells are used as input to the model. Conversely, in single cell approach, measurement results from single cell are used as input to the model.

When comparison of AI algorithms against non-AI algorithms is performed, same simulation assumptions are adopted for non-AI algorithms, which could be sample and hold for intra-frequency temporal domain prediction and pathloss offset-based algorithm for frequency domain prediction. Other simple models e.g. ARIMA(Autoregressive Integrated Moving Average) can be also considered. In sample and hold, the actual measurement result of the last time instance in OW is held for PW.

Simulation assumptions collected in the table 5.1-1 are for FR1 and FR2:

Table 5.1-1: Simulation assumptions of FR1 and FR2

|  |  |  |
| --- | --- | --- |
| Parameters | Value for FR1 | Value for FR2 |
| Frequency Range | FR1@{4GHz,30KHz} as central frequency for intra-frequency scenarioFR1@{2GHz, 15/30KHz} as another frequency for inter-frequency scenario | FR2 @ 30 GHz; SCS: 120 kHz |
| Deployment | 2-tier model with wrap-around (7 sites, 3 sectors/cells per site) | 2-tier model with wrap-around (7 sites, 3 sectors/cells per site) |
| Channel model | UMa With distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901, fast fading and optional LOSsoft; Inter-frequency correlation model is optional.without UErotation,Oxygen absorption, Time-varying Doppler shift, Explicit ground reflection model and blockage. | UMiWith distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901, fast fading and optional LOSsoft;without UE rotation,Oxygen absorption, Time-varying Doppler shift, Explicit ground reflection model and blockage |
| System BW | 20MHz | 80MHz |
| UE speed | 30,60,90 km/h for study targeting measurement reduction60,90,120 km/h for study targeting HO performance improvement | 30,60,90 km/h for study targeting measurement reduction60,90,120 km/h for study targeting HO performance improvement |
| UE distribution | 100% outdoor | 100% outdoor |
| BS Antenna Configuration | Companies need to report which option(s) are used between- 32 ports: (8,8,2,1,1,2,8), (dH,dV) = (0.5, 0.8)λ- 16 ports: (8,4,2,1,1,2,4), (dH,dV) = (0.5, 0.8)λ1,2 or 4 TX beams are assumed. | Antenna setup and port layouts at gNB: (4, 8, 2, 1, 1, 1, 1), (dV, dH) = (0.5, 0.5) λ8,16 or 32 TX beams are assumed |
| BS Antenna radiation pattern | 3-sector antenna radiation pattern, 8 dBi | TR 38.802 Table A.2.1-6, |
| UE Antenna Configuration | 4RX: (1,2,2,1,1,1,2), (dH,dV) = (0.5, 0.5)λ for (rank 1-4)2RX: (1,1,2,1,1,1,1), (dH,dV) = (0.5, 0.5)λ for (rank 1,2)1RX beam is assumed | Antenna setup and port layouts at UE: (1, 4, 2, 1, 2, 1, 1), 2 panels (left, right)4RX beams are assumed |
| UE Antenna radiation pattern | Omni-direction | TR 38.802 Table A.2.1-8,  |
| BS Tx Power | 44dBm  | 40 dBm (baseline)Other values (e.g., 34 dBm) not precluded |
| Maximum UE Tx Power | 23dBm | 23 dBm |
| BS receiver Noise Figure | 5dB | 7 dB |
| UE receiver Noise Figure | 9dB | 10 dB |
| Inter site distance | 500m | 200 m |
| BS Antenna height | 25m | 10m |
| UE Antenna height | 1.5m | 1.5 m |
| Spatial consistency | companies report one of the spatial consistency procedures: - Procedure A in TR38.901- Procedure B in TR38.901 | companies report one of the spatial consistency procedures: - Procedure A in TR38.901- Procedure B in TR38.901 |
| UE trajectory model | 3 options in 38.843 section 6.3.1 | 3 options in 38.843 section 6.3.1 |
| UE trajectory boundary processing model | Companies report which of the following models they used:wrap-around model, circle-bouncing model,boundary-terminated model | Companies report which of the following models they used:wrap-around model, circle-bouncing model,boundary-terminated model |
| Sampling period | 40ms | 80ms |

## 5.2 RRM measurement prediction

### 5.2.1 Evaluation methodology, metrics and assumptions

#### 5.2.1.1 RRM measurement prediction

Measurement prediction accuracy for cell-level RRM measurement prediction is defined as average L3 RSRP difference between predicted L3 filtered cell-level measurement result and ground truth L3 filtered cell-level measurement result of the same cell for all RRM sub-use cases.

Measurement reduction rate for intra-frequency scenario is defined in the temporal domain (called MRRT) by assuming same length of measurement time instances and in the spatial domain respectively (called MRRS):

MRRT = skipped measurement time instances / total measurement time instances

MRRS = skipped beams to be measured/ total beams to be measured

In intra-frequency temporal domain case A, continuous measurement results in PW are predicted by continuous historical measurement result(s) in OW. Then OW and PW slide forward with either sampling period(s) (with sliding L1/L3 filtering option) or measurement period(s) (with non-sliding L1/L3 filtering option), where measurement result(s) are actually measured before sliding. One example is illustrated in Figure 5.2.1.1-1:



Figure 5.2.1.1-1: Example of intra-frequency temporal domain case A

Intra-frequency temporal domain case A prediction is evaluated for the 2nd study goal for both FR1 and FR2 scenario.

In intra-frequency temporal domain case B, measurement results in PW are predicted by historical measurement result(s) in OW. Then OW and PW slide forward with either sampling period(s) (with sliding L1/L3 filtering option) or measurement period(s) (with non-sliding L1/L3 filtering option) and measurement result(s) in previous PW is/are skipped during window sliding. Example 1 and example 2 are illustrated in Figure 5.2.1.1-2 and Figure 5.2.1.1-3 respectively, between which example 2 is recommended as baseline for evaluation.

Note: The historical measurement results in OW are at least actual measurement results. Companies are free to report if they use predicted measurement results in OW as input of AI/ML model.



Figure 5.2.1.1-2: Skipping pattern example 1 of intra-frequency temporal domain case B



Figure 5.2.1.1-3: Skipping pattern example 2 of intra-frequency temporal domain case B

Intra-frequency intra-cell temporal domain case B prediction is evaluated for 1st study goal by predicting a sub set of measurement instances in temporal domain of the same cell for both FR1 and FR2 scenario. MRRT(s) should be aligned among companies without defining detailed skipping pattern. Both case A and case B are applicable for all RRM sub-use cases and focus on at least pure temporal domain.

Intra-frequency intra-cell spatial domain prediction is evaluated for the 1st study goal by measuring a sub set of configured SSB as input to the model to derive L3 filtered cell-level measurements for every time instance of the same cell. It is only evaluated for FR2 intra-frequency scenario and is applicable for RRM sub-use case 1 and 3. MRRS(s) should be aligned among companies without defining detailed pattern.

For both intra-frequency inter-cell prediction and FR1 to FR1 inter-frequency inter-cell prediction, no measurement is reduced in both temporal and spatial domain for cell to be measured. For FR1 to FR1 inter-frequency inter-cell prediction, focus on the case where cell to be measured and cell to be predicted are located in the same sector of either serving site or same neighbouring site. If inter-frequency correlation model is assumed, section 7.6.5 in [4] is taken as baseline for inter-frequency correlation model. FR1 to FR1 inter-frequency inter-cell prediction is applicable for all RRM sub-use cases.

Intra-frequency inter-cell prediction refers to neighbouring cell prediction based on measurements of either co-located or non-collocated serving cell or neighbouring cell.

The prioritization among evaluation scenarios is captured in table 5.2.1.1-1.

Table 5.2.1.1-1: Prioritization of evaluation scenarios

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| scenario number | Priority  | Evaluation scenario | Target study goal | Methodology |
| 1 | Low | FR1 to FR1 intra-frequency temporal domain case A | 2nd goal | intra-cell |
| 2 | High | FR1 to FR1 intra-frequency temporal domain case B | 1st goal | Intra-cell |
| 3 | High | FR1 to FR1 inter-frequency (frequency domain) | 1st goal | Inter-cell  |
| 4 | High | FR2 to FR2 intra-frequency temporal domain case A | 2nd goal | Intra-cell |
| 5 | Low | FR2 to FR2 intra-frequency temporal domain case B | 1st goal | Intra-cell |
| 6 | Middle | FR2 to FR2 intra-frequency spatial domain | 1st goal | Intra-cell |

Following RRC parameters are assumed for RRM measurement prediction:

Table 5.2.1.1-2

|  |  |
| --- | --- |
| L3 filtering parameter  | value |
| FR1 FilterCoefficient | 4 |
| FR2 FilterCoefficient | 4 |

Table 5.2.1.1-3

|  |  |
| --- | --- |
| Measurement period | value |
| FR1 to FR1 intra-frequency without gap | 200ms  |
| FR1 to FR1 inter-frequency with gap | 200ms |
| FR2 to FR2 intra-frequency without gap | 400ms  |

Table 5.2.1.1-4

|  |  |
| --- | --- |
| Consolidation parameter | value |
| nrofSS-BlocksToAverage for FR1 | 1 |
| nrofSS-BlocksToAverage for FR2 | 3 |
| absThreshSS-BlocksConsolidation for FR1 | -110dbm |
| absThreshSS-BlocksConsolidation for FR2 | -110dbm |

For FR1 inter-frequency prediction, Pearson correlation coefficient is used for correlation coefficient calculation.

#### 5.2.1.2 Generalization

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 and tested using the dataset with Configuration #B.

The detailed evaluation combinations of GC#1 and GC#2 on UE speed for both FR1 and FR2 are depicted in table 5.2.1.2-1.

Table 5.2.1.2-1: Evaluation combinations on UE speeds

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Training @Dataset: S1  | Training @Dataset: S2 | Training @Dataset: S3 | Inference @Dataset:S1 | Inference @Dataset:S2 | Inference @Dataset:S3 |
| Baseline | Yes  |  |  | Yes  |  |  |
| GC#1 |  | Yes |  | Yes |  |  |
| GC#1 |  |  | Yes | Yes |  |  |
| GC#2 | Yes | Yes | Yes | Yes |  |  |
| Baseline |  | Yes |  |  | Yes |  |
| GC#1 | Yes |  |  |  | Yes |  |
| GC#1 |  |  | Yes |  | Yes |  |
| GC#2 | Yes | Yes | Yes |  | Yes |  |
| Baseline |  |  | Yes |  |  | Yes  |
| GC#1 | Yes |  |  |  |  | Yes |
| GC#1 |  | Yes |  |  |  | Yes |
| GC#2 | Yes | Yes | Yes |  |  | Yes |

For FR1, the UE speed S1, S2 and S3 are 30 km/h, 60km/h and 90km/h. For FR2, the UE speed S1, S2 and S3 are 60 km/h, 90km/h and 120km/h.

The detailed evaluation combinations of GC#1 and GC#2 and the relevant set of cell configurations for FR1 or FR2 are depicted in able 5.2.1.2-2 and 5.2.1.2-3 respectively.

Table 5.2.1.2-2: Evaluation combinations on cell configuration

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Training @Dataset: CC1  | Training @Dataset: CC2 | Inference @Dataset:CC1 | Inference @Dataset:CC2 |
| Baseline | Yes  |  | Yes  |  |
| GC#1 |  | Yes | Yes |  |
| GC#2 | Yes | Yes | Yes |  |
| Baseline |  | Yes |  | Yes |
| GC#1 | Yes |  |  | Yes |
| GC#2 | Yes | Yes |  | Yes |

Table 5.2.1.2-3: Cell Configuration(CC) parameters

|  |  |  |
| --- | --- | --- |
| Parameter | Cell Configuration #1 | Cell Configuration #2 |
| Deployment scenario  | UMi | UMa |
| ISD | 200m | 500m |
| BS antenna height | 10m | 25m |
| BS Tx power | 40dBm | 44dBm |

The detailed evaluation combinations of GC#1 and GC#2 on FR1 inter-frequency prediction is depicted in table 5.2.1.2-4.

Table 5.2.1.2-4: Evaluation combinations on inter-frequency prediction

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Training @Dataset: 2GHz to 4GHz  | Training @Dataset: 4GHz to 2GHz | Inference @Dataset: 2GHz to4GHz | Inference @Dataset: 4GHz to 2GHz |
| Baseline | Yes  |  | Yes  |  |
| GC#1 |  | Yes | Yes |  |
| GC#2 | Yes | Yes | Yes |  |
| Baseline |  | Yes |  | Yes |
| GC#1 | Yes |  |  | Yes |
| GC#2 | Yes | Yes |  | Yes |

### 5.2.2 Evaluation results

In the evaluation, model complexity in number of model parameters, model complexity in ~~number of~~ model size, and computational complexity in FLOPs are used for AI/ML complexity analysis. Table 5.2.2-1 illustrates the complexity results for high-priority scenarios.

Table 5.2.2-1: AI/ML model complexity/computation complexity
used in the evaluations for AI/ML in RRM measurement prediction

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model complexity in number of model parameters | Model complexity in ~~number of~~ model size | Computational complexity (FLOPs) |
| FR1 to FR1 intra-frequency temporal domain case B | 16K to 1.51M majority reported less than 0.8M | 67Kbytes to 5.1Mbytes majority reported less than 1Mbytes | 0.12M to 23.86M majority reported less than 3M  |
| FR1 to FR1 inter-frequency (frequency domain) | 0.22K to 1.84M majority reported less than 0.33M | 2.1Kbytes to 7.3Mbytes majority reported less than 1Mbytes | 5K to 66M majority reported less than 3M |
| FR2 to FR2 intra-frequency temporal domain case A | 4.5k to 1.38M majority reported less than 0.7M | 69Kbytes to 10.8Mbytes majority reported less 0.74 Mbytes  | 20K to 33.1M majority reported less than 10M |

NOTE: Some simple models are feasible.

#### 5.2.2.1 RRM measurement prediction

##### 5.2.2.1.1 Basic performance for FR1 intra-frequency temporal domain case B

“RRM\_Scen 2” in the attached Spreadsheets presents the performance results for FR1 intra-frequency temporal domain case B.

A total of 15 companies provided their results for the scenario. Figures 5.2.2.1.1-1 and 5.2.2.1.1-2 compare the distributions of average L3-RSRP differences between AI/ML and non-AI approaches under MRRT = 50% and UE speed=30Km/h, for sliding and non-sliding filtering, respectively.



Figure 5.2.2.1.1-1: CDF for FR1 intra-frequency temporal domain case B with sliding L1/L3 filtering



Figure 5.2.2.1.1-2: CDF for FR1 intra-frequency temporal domain case B with non-sliding L1/L3 filtering

The detailed evaluation results of key parameters submitted by companies are summarized in Tables 5.2.2.1.1-1 and 5.2.2.1.1-2, corresponding to sliding filtering and non-sliding filtering, respectively.

In the performance results presented below:

* ‘Average’ refers to the averaged L3 cell-level RSRP difference across all the predicted instances within PW
* ‘Last’ refers to the average L3 cell-level RSRP difference of the last predicted instance within PW.

Table 5.2.2.1.1-1: Basic performance for FR1 intra-frequency temporal domain case B with sliding filtering

|  |  |  |  |
| --- | --- | --- | --- |
|  | UE speed | 30Km/h | 90Km/h |
| MRRT |  | AI | Non-AI | AI | Non-AI |
| 50% | Average [dB] | 0.06, 0.10, 0.10, 0.12, 0.26, 0.58, 0.66 | 0.10, 0.11, 0.13, 0.14, 0.38, 0.62, 0.70 | 0.08, 0.23, 0.45, 0.67, 1.23 | 0.20, 0.28, 0.63, 0.72, 1.21 |
| Last [dB] | 0.10, 0.10, 0.26, 1.02 | 0.11, 0.38, 1.23 | 0.23, 0.45, 0.89 | 0.28, 0.63, 1.31 |
| 66% | Average [dB] | 0.14, 0.38, 1.20 | 0.22, 0.75, 1.40 | 0.71, 1.19 | 1.25, 1.56 |
| Last [dB] | 0.20, 0.40, 1.80 | 0.29, 0.82, 2.02 | 0.71, 1.80 | 1.37, 2.35 |
| 80% | Average [dB] | 0.24, 0.25, 0.28, 0.66 | 0.30, 0.34, 0.37, 0.98 | 0.38, 0.72, 1.10 | 0.48, 0.84, 1.71 |
| Last [dB] | 0.33, 0.41, 0.76 | 0.40, 0.59, 1.18 | 0.85, 1.27 | 0.98, 2.01 |

Table 5.2.2.1.1-2: Basic performance for FR1 intra-frequency temporal domain case B with non-sliding filtering

|  |  |  |  |
| --- | --- | --- | --- |
|  | UE speed | 30Km/h | 90Km/h |
| MRRT |  | AI | Non-AI | AI | Non-AI |
| 50% | Average [dB] | 0.01, 0.06, 0.21, 0.26, 0.33, 0.45, 0.58, 0.96 | 0.03, 0.11, 0.41, 0.54, 0.63, 0.84 | 0.08, 0.09, 0.30, 0.88, 0.88, 0.91, 1.93 | 0.06, 0.95, 0.99, 1.10, 2.04 |
| Last [dB] | 0.21, 0.26, 0.33, 0.45 | 0.54 | 0.30, 0.88, 0.88 | 0.95, 1.10 |
| 66% | Average [dB] | 0.09, 0.25, 0.41, 0.51, 1.93 | 0.05, 0.61, 1.86 | 0.06, 1.34, 1.34, 3.68 | 0.11, 3.98 |
| Last [dB] | 0.09, 0.27, 0.53, 0.60 | 0.07 | 0.08, 1.70, 1.70 | 0.16 |
| 80% | Average [dB] | 0.11, 1.28, 1.52 | 0.10, 1.28, 1.73 | 0.17, 1.96, 2.13, 3.22 | 0.23, 1.96, 3.54 |
| Last [dB] | 0.15, 2.31 | 0.17, 2.42 | 0.22, 3.06, 4.53 | 0.38, 5.70 |

##### 5.2.2.1.2 Basic performance for FR1 inter-frequency prediction

“RRM\_Scen3” in the attached Spreadsheets presents the performance results for FR1 inter-frequency prediction.

A total of 11 companies provided their results for the scenario, Figure 5.2.2.1.2-1 and Table 5.2.2.1.2-1 illustrates the evaluation results of cell-based and cluster-based AI/ML models, respectively.



Figure 5.2.2.1.2-1: CDF for FR1 inter-frequency prediction

Table 5.2.2.1.2-1: Basic performance for FR1 inter-frequency prediction

|  |  |
| --- | --- |
|  | Average L3 cell-level RSRP difference [dB] |
| Model type | AI | Non-AI |
| Cell-based | 0.11, 0.23, 0.28, 0.82, 0.99, 2.29, 3.61, 4.28 | 0.80, 2.21, 3.24, 4.13 |
| Cluster-based | 0.20, 0.24, 0.43, 0.60, 1.00, 1.40, 2.94, 3.50 |

##### 5.2.2.1.3 Basic performance for FR2 intra-frequency temporal domain case A

“RRM\_Scen4” in the attached Spreadsheets presents the performance results for FR2 intra-frequency temporal domain case A.

A total of 14 companies provided their results for the scenario. Figures 5.2.2.1.3-1 compares the distributions of average RSRP differences between AI/ML and non-AI approaches under Speed = 60Km/h for sliding filtering for all PWs.



Figure 5.2.2.1.3-1: CDF for FR2 intra-frequency temporal domain case A with sliding filtering

The detailed evaluation results of key parameters submitted by companies are summarized in Tables 5.2.2.1.3-1 and 5.2.2.1.3-2, corresponding to sliding filtering and non-sliding filtering, respectively.

In the performance results presented below:

* ‘Average’ refers to the average L3 cell-level RSRP difference
* ‘Last’ refers to the averaged L3 cell-level RSRP difference of the last predicted point within PW.

Table 5.2.2.1.3-1: Basic performance for FR2 intra-frequency temporal domain case A with sliding filtering

|  |  |  |  |
| --- | --- | --- | --- |
|  | UE speed | 60Km/h | 120Km/h |
| PW |  | AI | Non-AI | AI | Non-AI |
| [40, 200] ms | Average [dB] | 0.22, 0.25, 0.26, 0.41, 0.41, 0.61, 0.69, 0.75 | 0.50, 0.65, 1.44 | 0.27, 0.63, 0.67, 0.71, 0.81, 0.97, 1.00 | 0.58, 0.70, 1.42 |
| Last [dB] | 0.35, 0.41, 0.41, 0.58, 0.84, 1.49 | 0.91, 2.75 | 0.63, 0.67, 1.11, 1.39, 1.80 | 0.99, 2.68 |
| [240, 400] ms | Average [dB] | 0.05, 0.61, 0.74, 0.74, 0.77, 1.15, 1.18, 1.29 | 1.16, 1.37, 1.75, 1.75 | 0.19, 0.82, 0.86, 1.45, 1.67, 1.72 | 1.37, 1.55, 2.09, 2.55 |
| Last [dB] | 0.11, 0.95, 1.00, 1.25, 1.75, 1.90, 1.94, 2.20 | 1.93, 2.22, 2.83, 2.91 | 0.42, 1.08, 2.00, 2.37, 2.92, 3.19 | 2.33, 2.46, 3.39, 4.01 |
| [480, 1600] ms | Average [dB] | 0.17, 0.88, 1.61 | 1.52, 1.66, 3.80 | 0.59, 1.13, 2.35 | 2.01, 3.43 |
| Last [dB] | 0.47, 1.73, 1.94 | 2.89, 2.90, 4.09 | 1.54, 2.38, 2.70 | 3.57, 6.21 |

Table 5.2.2.1.3-2: Basic performance for FR2 intra-frequency temporal domain case A with non-sliding filtering

|  |  |  |  |
| --- | --- | --- | --- |
|  | UE speed | 60Km/h | 120Km/h |
| PW |  | AI | Non-AI | AI | Non-AI |
| {400, 800, 1200, 1600} ms | Average [dB] | 1.12, 1.70, 1.74, 5.16 | 4.60 | 1.50, 2.10, 2.79 | 4.60  |
| Last [dB] | 1.12, 2.00, 6.76 | 5.90 | 1.50, 2.70 | 5.90 |

##### 5.2.2.1.4 Summary of performance results for RRM measurement prediction

Some general trends are observed for RRM measurement predictions based on the simulations performed for scenarios 2, 3 and 4 mentioned in Table 5.2.1.1-1.

For both FR2 intra-frequency temporal domain case A and FR1 intra-frequency temporal domain case B predictions, the following observations are made:

* Higher UE speed correlates with decreased prediction accuracy;
* Longer PW length correlates with decreased prediction accuracy;
* The gain of cluster approach against single cell approach is not clear.

For FR2 intra-frequency temporal domain case A the following observations are made:

* Increasing the OW length can improve the prediction accuracy, especially when the OW length is relatively short. However, once the OW length exceeds a certain value, further increase of the OW length does not yield significant benefit;
* A majority of the companies observe that RRM sub-use case 2 demonstrates higher prediction accuracy than RRM sub-use case 1 and RRM sub-use case 3 at least with short PW length;
* AI algorithm can outperform sample and hold in terms of prediction accuracy. The gain improves with increase of UE speed and PW length within a certain window length;

For FR1 intra-frequency temporal domain case B the following observations are made:

* Increasing MRRT correlates with decreased prediction accuracy;
* Under the same MRRT setting, different measurement skipping patterns can result in different prediction accuracy;
* When PW is short, the performance difference between AI algorithm and sample-and-hold is not significant. However, when PW becomes larger, AI algorithm outperforms sample-and-hold;
* AI algorithm can outperform sample and hold in terms of predication accuracy. The gain is higher with increase of UE speed and MRRT.

For FR1 inter-frequency predictions in co-located scenario, the following observations are made:

* The prediction accuracy is comparable between higher-to-lower frequency and lower-to-higher frequency case;
* The UE speed has minor impact on the prediction accuracy;
* The higher the correlation coefficient is between two frequency layers, the higher the prediction accuracy;
* The cluster approach can improve the prediction accuracy compared to single cell approach.
* AI algorithm with cluster approach shows better performance compared to pathloss offset-based algorithm. But AI algorithm with single cell approach achieves limited gain compared to pathloss offset based algorithm without the help of neighbour cell measurement results.

NOTE 1: “Higher-to-lower frequency case refers to the scenario where measurements on a lower frequency (2GHz in the simulations) were predicted based on the actual measurement results on a higher frequency (4GHz in the simulations) and vice versa for lower-to-higher frequency case.

#### 5.2.2.2 Generalization

##### 5.2.2.2.1 Generalization performance for FR1 intra-frequency temporal domain case B

“RRM\_Scen2\_Gen” in the attached Spreadsheets presents the generalization performance results for FR1 intra-frequency temporal domain case B.



Figure 5.2.2.2.1-1 CDF for prediction accuracy loss for intra-frequency temporal domain case B

A total of 7 companies provided their results for the scenario, Tables 5.2.2.2.1-1 and 5.2.2.2.1-2 illustrate the generalization performance across different UE speeds and across different cell configurations, respectively. Figure 5.2.2.2.1-1 illustrate the result for UE speed=30Km/h in Tables 5.2.2.2.1-1.

In the performance results presented below:

* ‘GC#1 - baseline’ refers to the prediction accuracy loss in terms of average L3 cell-level RSRP difference when comparing the results obtained using GC#1 to the baseline results
* ‘GC#2 - baseline’ refers to the prediction accuracy loss in terms of average L3 cell-level RSRP difference when comparing the results obtained using GC#2 to the baseline results

NOTE: A negative value indicates that GC performs better than the baseline, while a positive value indicates the opposite. The principle applies to all generalization tables.

Table 5.2.2.2.1-1: Generalization performance across different UE speeds for FR1 intra-frequency temporal domain case B with MRRT=50%

|  |  |  |
| --- | --- | --- |
| Testing dataset (UE speed) \ Accuracy loss | GC#1 – baseline [dB] | GC#2 - baseline [dB] |
| 30km/h | -0.037, -0.001, 0.002, 0.010, 0.020, 0.241 | -0.100, -0.056, -0.040, -0.002, 0.001, 0.003, 0.007, 0.044, 0.100 |
| 60km/h | -0.228, -0.012, -0.002, -0.001, 0.009 | -0.170, -0.123, -0.017, -0.001, 0.000, 0.004, 0.007, 0.023 |
| 90km/h | -0.422, 0, 0.004, 0.008, 0.016, 0.018 | -0.173, -0.080, -0.005, -0.002, -0.001, 0.000, 0.002, 0.010, 0.073 |

Table 5.2.2.2.1-2: Generalization performance across different cell configurations for FR1 intra-frequency temporal domain case B

|  |  |  |
| --- | --- | --- |
| Testing dataset \ Accuracy loss | GC#1 - baseline [dB] | GC#2 - baseline [dB] |
| Cell Configuration #1 | 0.003, 0.010, 0.010, 0.019, 0.023, 0.047 | -0.030, -0.009, -0.002, 0.000, 0.001, 0.002 |
| Cell Configuration #2 | 0.010, 0.010, 0.010, 0.020, 0.040, 0.074 | -0.031, -0.001, 0.000, 0.004, 0.005, 0.010 |

##### 5.2.2.2.2 Generalization performance for FR1 inter-frequency prediction

“RRM\_Scen3\_Gen” in the attached Spreadsheets presents the generalization performance results forFR1 inter-frequency prediction.



Figure 5.2.2.2.2-1 CDF for prediction accuracy loss of FR1 inter-frequency prediction

A total of 7 companies provided their results for the scenario, Table 5.2.2.2.2-1 illustrates the generalization performance across different frequency prediction directions. Figure 5.2.2.2.2-1 illustrates the 2G to 4G case in Table 5.2.2.2.2-1

In the performance results presented below:

* ‘GC#1 - baseline’ refers to the prediction accuracy loss in terms of average L3 cell-level RSRP difference when comparing the results obtained using GC#1 to the baseline results
* ‘GC#2 - baseline’ refers to the prediction accuracy loss in terms of average L3 cell-level RSRP difference when comparing the results obtained using GC#2 to the baseline results
* ‘2GHz -> 4GHz’ indicates that the model uses measurement results of 2 GHz as input to predict the corresponding measurement results at 4 GHz.
* ‘4GHz -> 2GHz’ indicates that the model uses measurement results of 4 GHz as input to predict the corresponding measurement results at 2 GHz.

Table 5.2.2.2.2-1: Generalization performance across different frequency prediction directions for FR1 inter-frequency prediction

|  |  |  |
| --- | --- | --- |
| Testing dataset \ Accuracy loss | GC#1 - baseline [dB] | GC#2 - baseline [dB] |
| 2GHz -> 4GHz | 0.010, 0.136, 1.509, 5.680, 10.320, 10.331, 16.838 | 0, 0.040, 0.057, 0.090, 1.021, 1.031, 1.811 |
| 4GHz -> 2GHz | 0.010, 0.194, 1.244, 5.440, 9.912, 10.950, 15.190, | 0, 0.030, 0.030, 0.055, 0.560, 0.989, 1.095 |

##### 5.2.2.2.3 Generalization performance for FR2 intra-frequency temporal domain case A

“RRM\_Scen4\_Gen” in the attached Spreadsheets presents the generalization performance results forFR2 intra-frequency temporal domain case A.



Figure 5.2.2.2.3-1 CDF for prediction accuracy loss of intra-frequency temporal domain case A

A total of 11 companies provided their results for the scenario, Tables 5.2.2.2.3-1 and 5.2.2.2.3-2 illustrate the generalization performance across different UE speeds and across different cell configurations, respectively. Figure 5.2.2.2.3-1 illustrates the case, i.e. UE speed=60Km/h, in Tables 5.2.2.2.3-1.

In the performance results presented below:

* ‘GC#1 - baseline’ refers to the accuracy loss in terms of average L3 cell-level RSRP difference when comparing the results obtained using GC#1 to the baseline results
* ‘GC#2 - baseline’ refers to the accuracy loss in terms of average L3 cell-level RSRP difference when comparing the results obtained using GC#2 to the baseline results

Table 5.2.2.2.3-1: Generalization performance across different UE speeds for FR2 intra-frequency temporal domain case A

|  |  |  |
| --- | --- | --- |
| Testing dataset (UE speed) \ Accuracy loss | GC#1 - baseline [dB] | GC#2 - baseline [dB] |
| 30km/h | 0.007, 0.860 | 0.010, 0.020 |
| 60km/h | -0.760, -0.001, 0.015, 0.020, 0.021, 0.425, 2.513 | -0.290, -0.064, -0.020, -0.003, -0.001, 0.018, 0.030, 0.145, 1.671 |
| 90km/h | -1.200, -0.374, 0.002, 0.003, 0.014, 2.184 | -0.250, -0.060, -0.030, -0.007, 0.007, 0.013, 0.165, 0.698 |
| 120km/h | -0.582, -0.007, 0.009, 0.010, 0.037, 0.050, 1.754 | -0.383, -0.340, -0.054, -0.030, 0.020, 0.024, 0.024, 0.036, 0.150 |

Table 5.2.2.2.3-1: Generalization performance across different cell configurations for FR2 intra-frequency temporal domain case A

|  |  |  |
| --- | --- | --- |
| Testing dataset \ Accuracy loss | GC#1 - baseline [dB] | GC#2 - baseline [dB] |
| Cell Configuration #1 | 0.050, 0.060 | 0.010, 0.024 |
| Cell Configuration #2 | 0.026, 0.050 | -0.011, 0.010 |

##### 5.2.2.2.4 Summary of performance results for generalization of RRM measurement prediction

For generalization over UE speeds, the following observations are made:

* Generalization performs well across all UE speeds in general;
* GC#2 slightly improves the prediction accuracy compared to GC#1;
* GC#2 offers comparable prediction accuracy as the baseline case for the same data set size;
* For GC#1, the smaller the UE speed difference is between training data set and inference data set, the closer prediction accuracy is to the baseline case.

For generalization over frequency domain prediction, the following observations are made:

* GC#2 always outperforms GC#1, and its prediction accuracy is close to the baseline case;
* Feeding the AI/ML model with the knowledge about the input & output frequency helps to improve prediction accuracy of GC#2;
* GC#1 suffers from significant performance loss without any pre-processing based on the information e.g. path loss difference.

For generalization over cell configurations for intra-frequency temporal domain case A in FR2 and case B in FR1, the following observations are made:

* Model is generalizable over cell configurations with different deployment scenarios (i.e., UMi and UMa);
* GC#2 slightly improves the prediction accuracy compared to GC#1, and its prediction accuracy is close to the baseline;
* The model trained in scenario with UMi channel model while tested in scenario with UMa channel model shows better performance than the other way around.

## 5.3 Measurement event prediction

### 5.3.1 Evaluation methodology, metrics and assumptions

The performance metric F1 score is defined as following:

F1 score = 2\*Precision\*Recall/(Precision + Recall)

Where:

Precision = n3/(n1+n3)

Recall =n3/(n2+n3)

For indirect prediction, the counter n1, n2 and n3 in the formula are defined as following:

* Counter n3(true event prediction): it increases by 1 when a ground-truth event occurs around a predicted event with ETD, whose range is [0, maximum ETD] or vice versa;
* Counter n1(false event detection): it increases by 1 when no ground-truth event occurs around a predicted event with ETD, whose range is [0, maximum ETD];
* Counter n2(missed event detection): it increases by 1 when no event is predicted around a ground-truth event with ETD, whose range is [0, maximum ETD].

The ETD, i.e. timing difference between ground-truth event and predicted event, is illustrated in Figure 5.3.1-1:



Figure 5.3.1-1: illustration of ETD

As illustrated in Figure 5.3.1-1, only if the ETD between a predicted event and a ground-truth event e.g. ground-truth event 2 is less than or equal to maximum ETD, the ETD can still be tolerated. Otherwise, both false event and missed event are detected.

For direct prediction, the counter n1, n2 and n3 in the formula are defined as following:

* Counter n3 (true event prediction): it increases by 1 when a ground-truth event occurs within the occurrence window of predicted event whose possibility is higher than a predefined threshold;
* Counter n1 (false event detection): it increases by 1 when no ground-truth event occurs within the occurrence window of predicted event whose possibility is higher than a predefined threshold;
* Counter n2 (missed event detection): it increases by 1 when a ground-truth event occurs, but it doesn’t fall in the occurrence window of any predicted event whose possibility is higher than a predefined threshold.

For direct prediction method, a measurement event could be predicted within an occurrence window starting from current time instance, i.e. t0, and future time instance t1 with a probability as illustrated in Figure 5.3.1-1.



Figure 5.3.1-2: occurrence window of direct prediction method

For measurement event prediction based on intra-frequency temporal domain case A or case B, the simulation assumptions for intra-frequency temporal domain case A or case B in Table 5.2.1-1, Table 5.2.1-2, Table 5.2.1-3 and Table 5.2.1-4 are reused respectively. On top of that, following additional simulation assumptions are used for measurement event prediction based on intra-frequency temporal domain case A in Table 5.3.1-1 and temporal domain case B in Table 5.3.1-2 respectively:

Table 5.3.1-1: Additional simulation assumptions for measurement event prediction based on intra-frequency temporal domain case A

|  |  |  |
| --- | --- | --- |
| 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) | N/A | Up to implementation |
| PW length (ms,\*\*) |  320 | Open for more values |
| Max ETD (ms, \*) | 80 | Open for more values |

Table 5.3.1-2: Additional simulation assumptions for measurement event prediction based on intra-frequency temporal domain case B

|  |  |  |
| --- | --- | --- |
| 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) | N/A | Up to implementation |
| PW length (ms,\*\*) | 200 (non-sliding)40 (sliding) | Open for more values |
| Max ETD (ms,\*) | 80 | Open for more values |
| MRRT | 50% | Open for more values |

\*: This parameter is only applicable for indirect prediction

\*\*: For direct prediction, PW length means the length of occurrence window. And for FR1 only baseline 200ms is applicable.

### 5.3.2 Evaluation results

#### 5.3.2.1 Performance of measurement event prediction based on FR2 intra-frequency temporal domain case A

“ME\_Indirect\_CaseA” and “ME\_Direct\_CaseA” in the attached Spreadsheets present the intermediate performance results for indirect and direct measurement event prediction based on FR2 intra-frequency temporal domain case A, respectively.



Figure 5.3.2.1-1 CDF for F1 score of measurement event prediction based on intra-frequency temporal domain case A

For measurement event prediction based on intra-frequency temporal domain case A, a total of 10 companies provided their evaluation results for F1 score, as illustrated in Table 5.3.2.1-1 and Figure 5.3.2.1-1.

Table 5.3.2.1-1: F1 score for indirect and direct measurement event predictions based on FR2 intra-frequency temporal domain case A

|  |  |  |
| --- | --- | --- |
| Metrics \ Methods | Indirect prediction | Direct prediction |
| F1 score | 0.59, 0.87, 0.92, 0.92, 0.95, 0.97, 0.98, 0.99 | 0.85, 0.92, 0.95, 0.96 |

#### 5.3.2.2 Performance of measurement event prediction based on FR1 intra-frequency temporal domain case B

“ME\_Indirect\_CaseB” in the attached Spreadsheets presents the intermediate performance results for indirect measurement event prediction based on FR1 intra-frequency temporal domain case B.



**Figure 5.3.2.2-1 CDF for F1 score of indirect measurement event prediction based on intra-frequency temporal domain case B**

For indirect measurement event prediction based on FR1 intra-frequency temporal domain case B, a total of 10 companies provided their evaluation results for F1 score, as illustrated in Table 5.3.2.2-1 and Figure 5.3.2.2-1.

Table 5.3.2.2-1: F1 score for measurement event prediction based on FR1 intra-frequency temporal domain case B

|  |  |  |
| --- | --- | --- |
| MRRT  | =50% | >50% |
| F1 score | 0.73, 0.88, 0.95, 0.96, 0.96, 0.97, 0.99, 0.99 | 0.24, 0.88, 0.94 |

#### 5.3.2.3 Summary of performance results for measurement event prediction

For indirect measurement event prediction based on FR2 intra-frequency temporal domain case A, the following observations are made:

* Most of the simulation results show that the F1 score is very good;
* F1 score is higher for shorter TTT values .

For indirect measurement event prediction based on FR1 intra-frequency temporal domain case B, the following observations are made:

* Very good F1 score can be achieved, which depends on filtering approach or PW length;
* Good F1 score can be achieved with small PW length;
* Higher MRRT value correlates with decreased F1 score.

F1 score for direct measurement event prediction is very good based on the simulation results by assuming 50% probability threshold.

## 5.4 RLF prediction

### 5.4.1 Evaluation methodology, metrics and assumptions

The metrics defined in section 5.3.1 including F1 score, Precision, Recall and related counter n1,n2 and n3 are reused for RLF prediction also.

Additional simulation assumptions on top of those in table 5.1-1 are listed in table 5.4.1-1:

Table 5.4.1-1

|  |  |
| --- | --- |
| Parameter | Value |
| Qin threshold | -6dB |
| Qout threshold | -8dB |
| Sample rate (TIndication\_interval) | 20ms (FR2)/40ms(FR1)  |
| Qin evaluation period | 100ms |
| Qout evaluation period | 200ms |
| T310 | 1000ms |
| N310 | 1 |
| N311 | 1 |
| Max ETD (ms, \*) | 80ms |
| PW length (ms, note2) | 400(FR1),400(FR2) |
| OW length (ms, \*) | Up to implementation |

\*: This parameter is only applicable for indirect prediction

\*: For direct prediction, PW length means the length of occurrence window, which is illustrated in Figure 5.3.1-2.

To simulate inference across cells, following assumptions are made for inference model:

* It is assumed that all cells are fully loaded for interference modelling and no resource scheduler is needed;
* Interference in simulation comes from co-site cells and surrounding 6 sites of serving cell, i.e., interference comes from 20 cells as illustrated in Figure 5.4.1-1;
* The beam with highest L1 RSRP of the serving cell is taken as serving beam, which is taken as the serving signal of RLM. And the beam transmission pattern is synchronized across the site/cells, i.e at any given time the transmitted beam index is the same across the site/cells.



Figure 5.4.1-1: Interference model

In Figure 5.4.1-1, cells in site1 are surrounded by cells in 2nd tier sites. Cells in the rest sites are surrounded by cells in 2nd tier sites and wrap rounded sites. Taking cells in site 6 example, they are surrounded by site 1,2b,3b,4a,5,7, where site 2b,3b and 4a are wrap rounded sites. The alternative solution is to set up 3 tier sites.

## 5.5 System level simulation

### 5.5.1 Evaluation methodology, metrics and assumptions

HOF model defined in section 5.2.1.3 of TR36.839 [7] is reused for SLS. The metric for SLS is HOF rate ,total number of handover attempts per UE per second and total number of handover failures per UE per second, which are defined in section 5.2.1.3 and section 5.4.2 of TR 38.839 [7] respectively. They are cited here:

The handover failure rate is defined as: Handover failure rate = (number of handover failures) / (Total number of handover attempts).

The total number of handover attempts is defined as: Total number of handover attempts = number of handover failures + number of successful handovers.

The total number of successful handovers per UE per second is defined as the total number of successful handovers averaged over the total travel time of all the simulated UEs

The total number of handover failures per UE per second is defined as the total number of handover failures averaged over the total travel time of all the simulated UEs

SLS is performed based on measurement event prediction defined in section 5.3. The simulation assumptions defined in section 5.3.1 are reused. The inference model defined in section 5.4.1 is reused also.

The handover model is defined to facilitate SLS, where measurement event is predicted based on either intra-frequency temporal domain case A or intra-frequency temporal domain case B. For both cases, network starts with 40ms handover preparation once a predicted measurement event is received. A handover command will be transmitted at least after preparation is completed. After handover command, 40ms execution duration is assumed.

If measurement event is predicted based on intra-frequency temporal domain case A, there are two options w.r.t. how to decide on the time point to transmit handover command:

Option 1: Relying on legacy measurement event

Option 2: Relying on predicted measurement event



Figure 5.5.1-1: Handover model option 1

Option 1 is illustrated in Figure 5.5.1-1. At current time i.e. t0 measurement event e.g. A3 event is predicted at some point of time in future. Network will not transmit handover command until a real measurement event is reported for the same neighbouring cell. In this way, the main benefit of this option is to save handover preparation time.



Figure 5.5.1-2: Handover model option 2

Option 2 is illustrated in Figure 5.5.1-2. At current time, i.e. t0, measurement event e.g. A3 event is predicted @ future time t1. Network transmits handover command when entry condition of the predicted measurement event is met based on actual measurement result @ t2 unless t2 is earlier than handover preparation phase. In later case, network transmits handover command immediately after handover preparation phase. In this way, not only handover preparation could be saved but also handover can be executed earlier.

If measurement event is predicted based on intra-frequency temporal domain case B, there is option 3 w.r.t. how to decide on the time point to transmit handover command:

 

Figure 5.5.1-3: Handover model option 3

Option 3 is illustrated in Figure 5.5.1-3. Once a predicted measurement event e.g. A3 event is received network can transmit handover command immediately after handover preparation is completed. UE will report predicted measurement event at the time instance it is to be triggered.

### 5.5.2 Evaluation results

#### 5.5.2.1 SLS Performance of measurement event prediction based on FR2 intra-frequency temporal domain case A

“ME\_Indirect\_CaseA” and “ME\_Direct\_CaseA” in the attached Spreadsheets present the SLS performance results for indirect and direct measurement event prediction based on FR2 intra-frequency temporal domain case A, respectively. Baseline in this section refers to HO performance of existing L3 handover procedure.



Figure 5.5.2.1-1 CDF for HOF rate difference based on FR2 intra-frequency temporal domain case A

A total of 7 companies provided their results for the scenario, Table 5.5.2.1-1 illustrates the SLS performance for both indirect and direct measurement event predictions. The SLS performance metrics include HO failure rate, total number of HOF per UE per second, and total number of HO attempts per UE per second. Figure 5.5.2.1-1 illustrates the HOF rate in Table 5.5.2.1-1.

In the performance results presented below:

* ‘(Indirect & option 1)- Baseline’ indicates the difference in the given metrics for indirect measurement event prediction when using handover model option 1, compared to the baseline.
* ‘(Indirect & option 2)-Baseline’ indicates the difference in the given metrics for indirect measurement event prediction when using handover model option 2, compared to the baseline.
* ‘(Direct & option 2)-Baseline’ indicates the difference in the given metrics for direct measurement event prediction when using handover model option 2, compared to the baseline.

Table 5.5.2.1-1: SLS performance for indirect and direct measurement event predictions based on FR2 intra-frequency temporal domain case A

|  |  |  |  |
| --- | --- | --- | --- |
| Metrics \ Gains | (Indirect & option 1) -Baseline | (Indirect & option 2) -Baseline | (Direct & option 2) -Baseline |
| HOF rate (%) | 0.647, 0, -0.540, -0.870 | 3.200, 0.410, 0.031, -2.210, -4.760, -12.000, -51.460 | -9.54 |
| Total number of HOF per UE per second | 0.002, 0.000, 0.000, -0.002 | 0, -0.007, -0.007, -0.019, -0.020, -0.349 | -0.029 |
| Total number of HO attempts per UE per second | 0.014, 0.010, 0.000, -0.032 | 0.020, 0.010, 0, -0.007, -0.157, -0.175 | -0.03 |

NOTE: A negative value indicates that AI/ML performs better than the baseline, while a positive value indicates the opposite.

#### 5.5.2.2 SLS Performance of measurement event prediction based on FR1 intra-frequency temporal domain case B

“ME\_Indirect\_CaseB” in the attached Spreadsheets presents the SLS performance results for indirect measurement event prediction based on FR1 intra-frequency temporal domain case B.

A total of 2 companies provided their results for the scenario, Table 5.5.2.2-1 illustrates the SLS performance for indirect measurement event predictions. The SLS performance metrics include HO failure rate, total number of HOF per UE per second, and total number of HO attempts per UE per second.

In the performance results presented below:

* ‘(Indirect & option 3) – Baseline’ indicates the difference in the given metrics for indirect measurement event prediction when using handover model option 3, compared to the baseline.

Table 5.5.2.2-1: SLS performance for indirect measurement event prediction based on temporal domain case B

|  |  |
| --- | --- |
| Metrics \ Performance degradation | (Indirect & option 3) – Baseline |
| HOF rate (%) | -1.00, 0.29 |
| Total number of HOF per UE per second | 0, 0 |
| Total number of HO attempts per UE per second | -0.01, 0 |

NOTE: A negative value indicates that AI/ML performs better than the baseline, while a positive value indicates the opposite.

#### 5.5.2.3 Summary of SLS Performance

Compared to the existing L3 handover mechanism:

* AI algorithm (with indirect measurement event prediction) following handover model option 1 and option 2 performs better than baseline in terms of HOF rate and total number of HOF per UE per second;
* Majority companies show that AI algorithm (with indirect measurement event prediction) following handover model option 2 outperforms handover model option 1 when RRM prediction accuracy is good enough. A few companies show opposite observation due to the risk of too early handover in handover model option 2;Few companies shows that AI algorithm with direct measurement event prediction methodology can reduce the total number of HOF per UE per second in SLS based on FR2 intra-frequency temporal domain case A;
* AI algorithm following handover model option 3 as illustrated in Figure 5.5.1-3 with MRRT=50% has a minor or even no degradation in terms of HOF rate and total number of handover attempts.

# 6 Potential specification impact

## 6.1 LCM, protocol and procedure aspects

### 6.1.1 Overview

Only functionality-based LCM is considered, i.e. model-based LCM is not considered. Scenarios including intra-frequency temporal domain case A, intra-frequency temporal domain case B, intra-frequency spatial domain prediction and inter-frequency prediction are considered. Both L3 cell level prediction and L3 beam level prediction are considered.

RRM measurement prediction can be performed using Either UE-sided model or network-sided model. To support measurement event prediction, spec enhancements can be considered only for UE-sided model. How to predict measurement event using a network sided model is up to network implementation without spec impact.

RSRP is the baseline measurement quantity.

NOTE 1: Model transfer/delivery and data transfer for UE-sided model are not discussed in this study item.

NOTE 2: Spatial domain prediction across cells is up to network’s implementation

### 6.1.2 RRM measurement prediction

#### 6.1.2.1 UE-sided model

##### 6.1.2.1.1 Applicability reporting

UE can be configured with a full inference configuration and/or a partial inference configuration with inference parameters defined in section 6.1.2.1.2 in a *RRCReconfiguration* message. Upon transition to RRC\_IDLE or RRC\_INACTIVE state or upon RLF, UE follows existing behaviour defined in [2] on how to handle an inference configuration.

Upon receiving an inference configuration via *RRCReconfiguration* message, UE reports whether it is applicable or inapplicable in initial applicability report via *RRCReconfigurationComplete* message. If an inference configuration is inapplicable, UE-may include a flag to indicate its preference to release an inapplicable configuration . When UE indicates that an inference configuration is inapplicable, network is expected to release it i.e., autonomous release by UE is not supported. For an inapplicable full inference configuration, UE continues to perform the inference. And it is up to network implementation how to do with reported inference result after UE indicates the corresponding full inference configuration is inapplicable. If an inference configuration is applicable, UE applies the inference configuration without dynamic lower layer signalling.

Applicability can be updated via UAI. A flag in *OtherConfig* is introduced to indicate whether applicability reporting via UAI is enabled or not. When an inference configuration becomes inapplicable UE shall report its inapplicability via UAI.

No prohibit timer need be introduced for applicability reporting.

##### 6.1.2.1.2 Inference configuration and report

Existing RRM measurement configuration and reporting framework in RRC layer is baseline for inference configuration and report. When a full inference configuration is received, UE shall maintain it until it is released by network explicitly.

Following inference parameters can be configured to UE for inference and assessing applicability:

* PW length for intra-frequency temporal domain case A
* Measured frequency carrier and predicted frequency carrier information for inter-frequency prediction
* Parameter for intra-frequency temporal domain case B to indicate the timing of network’s SSB configuration instead of timing for UE to perform or skip measurement.
* Optional list of cells for intra-frequency temporal domain case A, for which network expects inference report (if available)
* Optional Beam pattern e.g. to save SSB transmission for intra-frequency spatial domain prediction
* Optional associated ID

It is up to UE’s implementation to decide on model related choices including cluster-based vs single-cell-based approach, RRM sub-use cases and OW length.

NOTE1: MRRT or MRRS can be considered as inference parameter if it is required for defining performance requirement in RAN4

NOTE 1: The detailed design of associated ID will be figured out during WI phase

UE can be configured with periodic or event triggered reporting of predicted and/or actual measurement result(s). For each predicted cell in the *measurementReport* message:

* For intra-frequency temporal domain case A, one or more instances of predicted measurement result(s) in PW is reported,
* For intra-frequency temporal domain case B, the latest actual or predicted measurement result is reported
* For inter-frequency prediction, the latest predicted measurement result is reported

##### 6.1.2.1.3 Monitoring and management

Performance of UE-sided model can be monitored in either network side or UE side after UE is configured with monitoring configuration and inference configuration. A monitoring window can be configured, over which the performance monitoring metric can be calculated.

For network-sided monitoring, UE can report ground-truth measurement result and inference output to network. And it is up to network implementation to perform monitoring and make further management decision.

For UE-sided monitoring, UE can report performance result, i.e. RSRP difference, to network based on ground-truth measurement result and inference output. It is up to network implementation to make management decision based on received performance result.

For UE-sided monitoring, it can be considered for UE to make management decision based on network’s configuration and report the decision to network instead of performance result.

##### 6.1.2.1.4 Data collection for offline training

UE can request data collection configuration via UAI message. The request can contain an indication on start or stop of data collection. And it is up to UE implementation when to send the request. The network can configure whether UE is allowed to initiate request.

The network can provide or release the data collection configuration at any point of time regardless of UE’s request. And network can decide when to start or stop the data collection. Data collection related configuration (e.g. MO(s) configured for legacy RRM measurement) and associated ID(s)(if needed) can be included in data collection configuration.

There are two options to decide on frequency(s) for data collection measurement:

* Option 1: Network can configure a set of candidate frequencies at first. Then UE can indicate a preference within the set of candidate frequencies to network.
* Option 2: UE can indicate preferred frequency(s) directly without set of candidate frequency(s) from network, which is under network control.

In both options, information other than frequency may be needed. And in option 1, candidate configurations should not be a list of full measurement configuration.

#### 6.1.2.2 Network-sided model

##### 6.1.2.2.1 Inference input reporting

For inference operation of network-sided model, the existing RRM measurement configuration and reporting framework in RRC layer can be reused. In addition, measurement result per cell at multiple time instances can be reported within one measurement report message for intra-frequency temporal domain case A sub-case 2.

NOTE 1: L1-filtered beam-level RSRP can be reported by configuring the corresponding *FilterCoefficient* to zero, if any

##### 6.1.2.2.2 Monitoring and management

For performance monitoring of network-sided model, the legacy RRM measurement configuration and reporting framework in RRC layer can be reused i.e. no spec impact is identified.

And UE will not be informed about any network-sided functionality management decision.

NOTE 1: UE awareness and preference when legacy measurement result is reported for network-sided model can be discussed in WI phase

##### 6.1.2.2.3 Data collection for offline training

Based on existing RRM measurement configuration framework UE can be configured to log L3 cell/beam level measurement result, L1-filtered beam level measurement result, cell identity information and timing information. For serving cell the cell identity information could be CGI if it is available or PCI + ARFCN otherwise. For neighbouring cell, the cell identity is PCI+ARFCN.

NOTE 1: Whether existing *measConfig* structure is reused or separate logging configuration is introduced will be discussed in WI phase.

UE performs logging periodically. If UE is configured with a L3 measurement event, it starts logging only when the L3 measurement event is fulfilled. UE can be configured to send availability indication of logged data via UAI or *RRCReconfigurationComplete* message (for HO case) when full buffer or preconfigured buffer threshold is reached or battery is low. The availability indication can indicate availability of logged data and/or triggering reason related to buffer and/or low power state.

NOTE 2: Whether condition of full buffer or buffer threshold is per use case or per UE can be discussed in WI phase.

Upon receiving *UEInformationRequest* message from network, UE sends the logged data via *UEInformationResponse* message.

UE keeps the logged data during handover procedure unless explicitly indicated by the network to release it. Upon transiting to RRC\_IDLE or RRC\_INACTIVE state UE releases logged data. However, it is beneficial to keep logged data upon RLF.

NOTE 3: Whether keeping logged data upon RLF depends on whether a simple solution can be defined in WI phase.

NOTE 4: UE can perform data collection in IDLE/INACTIVE mode without any specification impact

### 6.1.3 Measurement event prediction

It is up to UE’s implementation to decide on choice between indirect and direct event prediction methodology.

On top of inference parameters captured in section 6.1.2.1.2, event-related parameters e.g. event type are part of inference configuration of measurement event prediction.

UE can be configured with event triggered reporting based on predicted and/or actual measurement result(s). As baseline event type A1~A6 can be predicted and reported. UE can report measurement results, predicted measurement event together with its timing related information. When UE is configured with intra-frequency temporal domain case B or inter-frequency prediction, UE reports measurement event by following existing procedure. This can be achieved without specification impact.

For indirect event prediction, RSRP difference can be used as performance monitoring metric. And there is no consensus on the feasibility of performance monitoring of direct event prediction.

NOTE 1: The spec impact captured in section 6.1.2.1 is applicable for measurement event prediction unless otherwise described explicitly in this section.

NOTE 2: A single framework is aimed for direct and indirect event prediction

NOTE 3: The feasibility of performance monitoring of direct event prediction should be concluded before proceeding with normative work.

## 6.2 Interoperability, testability and RRM requirements

### 6.2.1 RRM requirements for measurement prediction

#### 6.2.1.1 General

The impact on RRM requirements for L3 cell-level RSRP (i.e., Point C as defined in Figure 9.2.4-1 in TS 38.300) measurement prediction has been studied for scenarios 2, 3, and 4 as defined in Table 5.2.1.1-1 and the sub-use cases 1, 2, and 3 as defined in section 4.2.

Editor Note: The impact of L3 beam level (Point E as defined in Figure 9.2.4-1 in TS 38.300) measurement prediction has not been studied. However, it doesn’t preclude the discussion on L3 beam level measurement prediction in the work item phase.

Editor Note: From RAN4 perspective, it is not precluded that other scenarios in Table 5.2.1.1-1 will be discussed in work item.

#### 6.2.1.2 Potential RRM requirements

The requirement impact corresponding to Point A1 as defined in Figure 9.2.4-1 in TS 38.300 for measurement prediction is not to be studied.

Both absolute and relative accuracy of predicted L3-RSRP are considered as the candidate metrics for RRM measurement prediction use cases for intra-frequency scenario and inter-frequency scenarios.

The absolute accuracy of predicted L3-RSRP is defined as:

* Absolute accuracy of predicted L3-RSRP = reported predicted L3-RSRP – ground truth of L3-RSRP.

The relative accuracy of predicted L3-RSRP for intra-frequency prediction is defined as:

* Relative accuracy of predicted L3-RSRP = (reported predicted L3-RSRP of cell 1 – reported L3-RSRP of cell 2) – (ground truth of L3-RSRP of cell 1 – ground truth of L3-RSRP of cell 2),
	+ cell 1 and cell 2 are on the same frequency
	+ the reported L3-RSRP of cell 2 can be measured or predicted.
* Editor Note: The relative accuracy of predicted L3-RSRP for beam level measurements may be further discussed during WI phase depending upon RAN2 progress.

The relative accuracy of predicted L3-RSRP for inter-frequency prediction is defined as:

* Relative accuracy of predicted L3-RSRP = (reported predicted L3-RSRP of cell 1 – reported L3-RSRP of cell 2) – (ground truth of L3-RSRP of cell 1 – ground truth of L3-RSRP of cell 2),
	+ cell 2 is on a different frequency than cell 1 but in the same FR as cell 1
	+ the reported L3-RSRP of cell 2 can be measured or predicted.

Note: It is not precluded to update the definition based on further RAN2 progress in WI phase.

As a baseline, the ground truth of L3-RSRP in FR2 is defined as the reported L3-RSRP measurement result under sufficient high SNR.

The ground truth of L3-RSRP in FR1 is based on:

* Alt1: The transmitted or reception power
* Alt2: The reported measurement value under certain conditions.

For intra-frequency temporal domain prediction, the impact on prediction accuracy by UE speed, MRRT, and OW/PW length will be considered. The number of measurements performed in OW could be also considered.

For inter-frequency prediction scenario, the impact on prediction accuracy by following elements can be considered:

* Side condition of frequency prediction (e.g., EPRE difference)
* Cluster approach, e.g.,
	+ when measurement from single cell in one carrier frequency is used by the UE as an input to predict the RRM measurement for the intra-FR and co-located cell in another carrier frequency.
	+ When measurement from a group of cells in one carrier frequency is used by the UE as an input to predict the RRM measurement for the intra-FR and co-located cell in another carrier frequency.
* RAN4 may further discuss whether and how correlation coefficient can be considered in simulation assumption for RAN4 requirements.

For both intra-frequency and inter-frequency prediction, RAN4 thinks RF and baseband errors on the measurements would impact the prediction accuracy.

* To model the measurement error, the following can be used as a baseline:
	+ For BB error, use link level simulation to generate L3-RSRP difference as baseband error.
	+ For RF error model, use a truncated Gaussian distribution.

### 6.2.2 RRM requirements for measurement event prediction

#### 6.2.2.1 General

The impact on RRM requirements for both direct and indirect event prediction use cases as defined in section 5.3 has been studied.

#### 6.2.2.2 Potential RRM requirements

For indirect event prediction, the requirements for the predicted event triggered reporting including, but not limited to, delay and accuracy requirements and the performance metrics are considered. If the report includes the predicted RSRPs corresponding to the predicted event occurrence time, the absolute and/or relative accuracy requirement for the predicted RSRP will be defined.

### 6.2.3 Testability for RRM measurement prediction

#### 6.2.3.1 Testing goal

As a baseline, the testing goal is to verify whether the minimum performance of AI/ML functionality/feature can be achieved.

#### 6.2.3.2 Prediction consistency in time domain

To ensure prediction consistency in the time domain, it is considered how to model the different time-varying characteristics per cell/site due to moving UE trajectories. It is also considered how to incorporate controlled randomness and the extent of time-domain variation and correlation.

FR1 is prioritized during the discussion.

#### 6.2.3.3 Testing setup

For testing in FR1, conducted testing is considered as a baseline.

For the number of cells configured in the testing, two cells are considered including serving cell and another/target cell for intra-cell RRM measurement prediction/event prediction. In the test, the measurement and prediction are performed on the same cell.

Note: Whether more than 2 cells are needed for inter-cell RRM measurement prediction/event prediction will be decided in WI phase.

### 6.2.4 Interoperability

RAN4 does not identify any interoperability issue which impacts the feasibility.

### 6.2.5 Generalization

RAN4 to discuss the requirement and testing aspects related to the generalization in WI phase.

# 7 Conclusion

The study focuses on evaluation of benefit of AI mobility use cases, namely RRM measurement prediction and measurement event prediction. Another use case i.e. RLF prediction is studied without evaluation. The potential specification impact is also studied to enable RRM measurement prediction, measurement event prediction and relevant mobility procedure in RRC\_CONNECTED state within NR system.

During the study, FR1 intra-frequency temporal domain case B and FR1 inter-frequency prediction are chosen as representative scenarios to verify study goal1, i.e. measurement reduction. For FR1 intra-frequency temporal domain case B ,the simulation results captured in section 5.5.2.2 show that there is no considerable handover performance degradation compared with existing L3 handover procedure when measurement is reduced e.g. 50% in temporal domain. For inter-frequency prediction, in addition to reducing UE’s measurement efforts, the UE throughput can also be increased if measurement gap configuration can be avoided or relaxed.

FR2 intra-frequency temporal domain case A is chosen as a representative scenario to verify study goal2, i.e. to improve handover performance (the reduction of handover failure (HOF) rate, etc). The simulation results captured in section 5.5.2.1 indicate reduction of the HOF rate in most cases when the handover is executed based on predicted measurement event in advance. For other companies, the HOF rate is not changed significantly.

The simulation results for RRM measurement prediction captured in section 5.2.2.1 show that the AI algorithm outperforms non-AI (e.g. sample and hold) in terms of prediction accuracy, i.e. average difference between actual and predicted L3 cell level RSRP values for both intra-frequency temporal cases A and B and for inter-frequency prediction, especially for long prediction windows.

Furthermore, simulation results for generalization captured in 5.2.2.2 show that the AI models can generalize well across UE speeds and different cell configurations, especially when the training is performed using mixed data sets or inter-frequency prediction direction is indicated.

Limited simulation results are submitted for intra-cell spatial domain prediction and L3 beam level prediction.

Both cluster approach (where measurement results from more than one cells are used as input to the model) and single cell approach (where measurement results from single cell are used as input to the model) were used by different companies. Both approaches are valid implementations.

Specification impact for both UE sided model and network sided model are studied. The study focused on potential enhancements of LCM procedures including data collection for model training. The outcome of the study is captured in section 6.1 and 6.2. For UE sided model the specification impact is mainly due to the introduction of RRM measurement prediction, with limited additional specification impact for measurement event prediction. The main specification impact for network sided model is for data collection.

For RRM measurement prediction, L3 beam-level prediction is feasible, however there are concerns on UE complexity and uncertainty of impacts/evaluations in other WG ~~workload~~ ~~uncertainty~~ for UE sided model.

For network sided model, all scenarios and all RRM sub-cases are feasible based on existing specification. For intra-frequency temporal domain case A sub-case 2 enhancement is needed. For other cases there is no specification impact. However, they can be discussed in WI phase whether enhancement (i.e. multi-instances reporting of beam) is needed and justified.

Based on what is summarized above, ~~we~~ ~~recommend~~ RRM and measurement event prediction are recommended for normative work .And the following scenarios and/or sub-cases are recommended for normative work:

* For UE sided model (RRM and measurement event prediction), intra-frequency temporal domain case A, intra-frequency temporal domain case B and inter-frequency domain prediction for co-located case,
* For network sided model (RRM prediction), at least RRM sub-case 2 of intra-frequency temporal domain case A for inference input report and all scenarios and sub-cases for data collection.

Annex <A> Change history:

|  |
| --- |
| Change history |
| Date | Meeting | TDoc | CR | Rev | Cat | Subject/Comment | New version |
| 2024-06-11 | RAN2#126 | R2-2406096 |  |  |  | Endorsed skeleton | 0.02 |
| 2024-08-21 | RAN2#127 | R2-2406309 |  |  |  | Endorsed text proposal | 0.03 |
| 2024-10-14 | RAN2#127bis | R2-2409011 |  |  |  | Endorsed text proposal | 0.04 |
| 2024-11-18 | RAN2#128 | R2-2410186 |  |  |  | Endorsed text proposal | 0.05 |
| 2025-02-21 | RAN2#129 | R2-2500287 |  |  |  | Endorsed text proposal | 0.06 |
| 2025-04-08 | RAN2#129bis | R2-2501822 |  |  |  | Endorsed text proposal | 0.07 |
| 2025-05-23 | RAN2#130 | R2-2503541 |  |  |  | Endorsed text proposal | 0.08 |
| 2025-06-10 | RAN#108 | RP‑250971 |  |  |  | Endorsed text proposal | 1.0.0 |
| 2025-08-25 | RAN2#131 | R2-2505161 |  |  |  | Endorsed text proposal | 1.1.0 |