AUSTRALIAN ENERGY MARKET COMMISSION

FREQUENCY PERFORMANCE PAYMENTS ANALYSIS

CONSULTANT TASK 3 OF ERC0263 – PRIMARY FREQUENCY RESPONSE INCENTIVE ARRANGEMENTS

Hugh Bannister
Jabez Wilson
Dr Sami Aoude

19 May 2022

Final Report

IES
Intelligent Energy Systems

Executive Summary

Background

On 16th September 2021, the AEMC published a draft determination describing a new rule for enduring arrangements to support the control of power system frequency and to incentivise plant behaviour that reduces the overall cost of frequency control during normal operation. The draft rule seeks to improve on the current causer pays process in three main ways.

- by making the measurements and settlement concurrent
- by weighting the performance factors by DI to reflect different market conditions
- by providing a positive incentive for good performers

Stakeholder feedback has indicated that there was not enough detail or analysis of the proposed new arrangements given in the Draft Determination. The AEMC has acknowledged this feedback and has engaged IES to help investigate options for the incentives framework before coming to a final rule. This document reports on the results of that investigation. It is informed by the oversight of the project team with members from the AEMC, AEMO and the consultant. However, the content of this report is entirely the responsibility of the consultant.

Project Aim and Approach

The aim of the project is to assist the AEMC investigate a range of options for a performance-based incentive arrangement to support the control of power system frequency. The results of the IES analysis will inform the AEMC and its stakeholders and support the selection of a preferred arrangement. This report presents details the process undertaken and the results of this work. The broad process to be followed is as follows.

1. Initial screening

The first step is to conduct a qualitative assessment to eliminate options that are inconsistent with the assessment framework. We perform a similar analysis on combinations of options to develop an initial set of candidate combinations.

2. Short sample period assessment

In this step quantitative analysis is conducted on the initial set by testing it over a representative two weeks of 4-second data from the full sample period. A final set of the best-performing combinations (candidates) is then selected for more detailed analysis.

3. Long sample period assessment

In this step, the final candidate combinations are tested using data from the entire sample period of 11 months. It is envisaged that a maximum of four combinations (scenarios) will reach this stage.

The analysis will use historical SCADA and market data from 1 February 2021 to 26 December 2021 (Inclusive, rounded to whole weeks from the 31 December end date originally specified) which will give an initial indication of the effect of the proposed arrangements.

This is a technical report. It assumes that the reader is familiar with the concept of frequency control and the AEMC and AEMO consideration of this topic in recent years. The AEMC Draft Determination and Rule is the key reference.

Summary of findings

The project brief recognised that the procedure to implement the new Rule had a set of identifiable design elements, each of which could support a set of options to be examined. These elements are set out in the table below. A set of identifiers was defined and used to help organise the work program.

Name	Identifier
Performance Metric	PM
Reference Trajectory	RT
Classification	C
Unit Aggregation	UA
Residual Calculation	RC
Financial Weight	F
Scaling	S
Regulation Enablement Cost Distribution	EC

While the are many design elements in proposed arrangements, a small number turned out to be critical.

4. The choice of a performance metric. Options considered were

    - A raw Hz metric (measured at AGC rate – 4 seconds on mainland), intended to capture the requirement for Primary Frequency Response (PFR)
    - Statistical measures such as correlation between unit and frequency deviations
    - A smoothed Hz metric, reflecting the requirement for a sustained response (and somewhat similar to the AGC Frequency Indicator (FI) metric but simpler and more transparent)
    - A combined Hz metric that captures both the raw and smoothed components, in equal measure
    - Metrics based on summations of measured unit deviations:
        - Gross system error, being the absolute sum of deviations working to correct the residual error
        - Net system error, being the algebraic sum of all errors in the system, assumed to be driving frequency deviations

Investigations confirmed that a combined frequency-based metric that had two components that targeted both primary and secondary response was preferable for the following reasons:

- it is a direct measure of the requirement for good frequency response;
- it can be expressed in a formula based on system frequency, which can be measured and tracked locally by all market participants; and
- when applied, it gave results consistent with current unit and system performance.

Investigation revealed that metrics based on MW deviations were poorly correlated with frequency and could not be expected to incentivise good frequency control.

5. The choice of reference trajectory over a dispatch interval (DI). The options considered were:

   - a linear ramp from scheduled target in the energy market to scheduled target;
   - a linear ramp from an initial measurement to a scheduled target;
   - an energy market trajectory modified by Mandatory PFR performance settings;
   - a trajectory determined by the AGC base point; and
   - an energy market trajectory that includes the AGC trajectory.

   The target-to-target option was preferred for the following reasons:

   - it is a simple and transparent with no unit-specific characteristics other than those present in the energy market; and
   - it reflects the trajectory of fully conforming units in the absence of frequency deviations and AGC reg variation, and so provides a good reference.

- Alternatives considered but rejected were accounting for the MPFR variation and using the AGC base point from AGC. Issues associated with including the AGC reg trajectory were examined but left open for further consultation if desired. Our preference is not to include it as it adjusts to poor performance and is thus a poor reference.

6. The choice of a scaling approach: This was examined in some detail and options narrowed down but further detailed consideration is desirable before finally settling on an option. Options considered in detail in both short and long period analyses were:

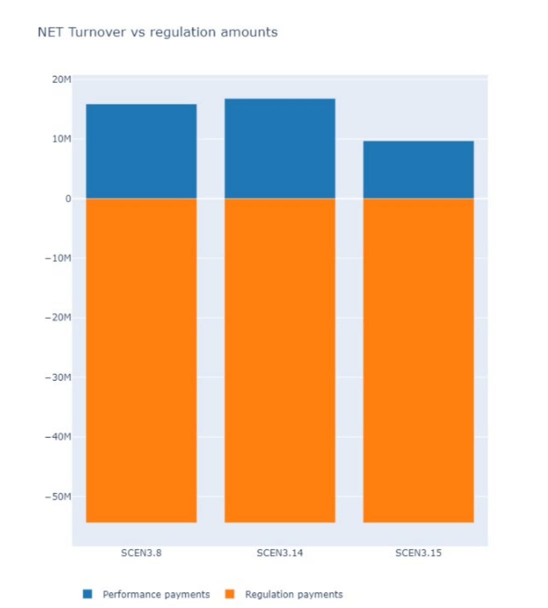
   - Scenario 3.8: Scaling by Gross System Error measured as the maximum deviation looking back over a DI. Other elements are standardised at preferred options.
   - Scenario 3.14: A frequency deviation pricing approach, where the deviation price is equal to the chosen Financial Weight when the frequency deviation crosses the Mandatory PFR dead band of 15 mHz. Other elements are standardised at preferred options.

- Scenario 3:15: Based on Scenario 3.8 but including unit trajectories determined by their AGC enablement instructions, as set by their AGC base points. The Gross System Error is adjusted down by the amounts provided by enabled units to reflect the fact that AGC reg enabled units have partially met the requirement for frequency control outside the incentive arrangement.

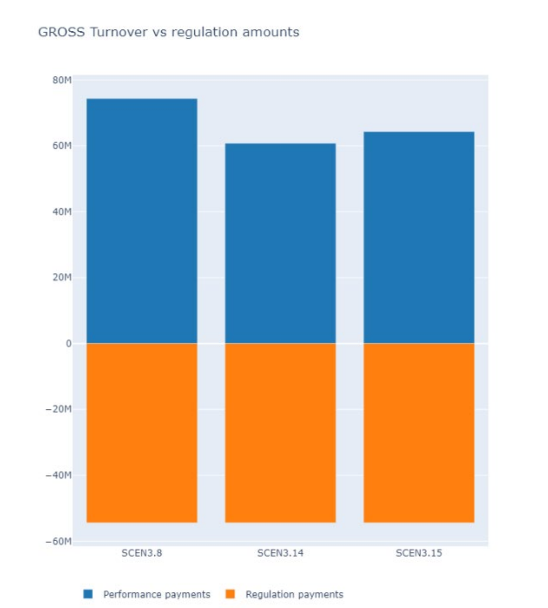
Summary of Final Scenario Results

Comparative settlement results for both performance and regulation for the three final scenarios are shown in the figure below. Raise and Lower Results are combined and the analysis was carried out over 11 months, from 1 February 2019 to 26 December, 2019. The Gross chart on the right shows turnover at the DI level, added up with positive and negative values preserved. The net chart on the left shows the effective bills, where the positive and negative settlements for a unit are netted out. Note the difference in scales: the regulation amounts are all identical/

Net Turnover for three Scaling Scenarios



Gross Turnover for three Scaling Scenarios



The merits of each of the approaches 3.8 and 3.14 need investigation in more detail, to see how the scaling works in different system behaviour scenarios. The 3.15 Scenario is provided for reference although we argue in the text that there are disadvantages to this approach.

The notable feature of these summary results is that the gross (DI level) performance payments are of the same order as the regulation payments but reduce to about a third or less of regulation costs when netted out over a billing period. This is due to the relatively poor correlation between deviations and the proposed frequency metric at present. This should change as participants begin to respond to the incentives.

We observe somewhat similar outcomes between 3.8 and 3.14, although 3.14 suffers somewhat less shrinkage when moving from gross to net. This suggests that 3.14 is less influenced by the current noise in the system, a property that is worthy of investigation at a detailed level during AEMO's consultation process.

It is notable that the net amount in 3.15 is significantly less than the gross, shrinking in a greater proportion than the other two. This reflects the greater influence of non-enabled units when the reg trajectory is included. While performing PFR, non-enabled units are currently not contributing significantly to the regulation requirement.

Another observation is that the gross turnover is not significantly less when AGC regulation unit contributions are largely removed. This suggests that the scaling MW is reduced only by a relatively small amount on average in this case, leaving the non-enabled units to share a similarly sized financial pot. Again, this points to a characteristic of the gross system error scaling approach which warrants more detailed investigation by AEMO.

Summary of Element Recommendations

These and other outcomes from the investigation as well as the preferred approach for each element of the procedure are summarised in more detail in the table following.

Summary of Recommendations

Element	Recommendation	Additional comment
Performance Metric	Combined frequency metric: Hz and smoothed Hz. Reasons: • Frequency performance is the most direct measure of good control. • A combined metric captures the need for immediate, fast acting response as well as the need to sustain that response even as frequency deviations return to zero.	• No dead band was applied. • Equal weighting of raw frequency and time filtered components, but subject to review based on performance. • No AGC-based time error correction but consider an additional smoothed linear frequency offset component based on time error only, with no AGC input.
Reference Trajectory	Linear ramp from scheduled target-to-target with AGC reg trajectory excluded. Reasons: • This trajectory reflects what a well-performing unit should do in the energy market. • Avoids incentive discontinuities at DI boundaries	• The inclusion of AGC reg is an option but we consider it would distort the incentive arrangements by upsetting the natural balance of performance factors. However, this option could be subject to a deeper investigation. • Using nominal MPFR trajectories would not reward units for the service provided. • Using AGC base points is inappropriate because they are adjusted to suit the performance of each unit and are intended to deliver the scheduled linear ramp.
Classification	Separate Raise and Lower payments and base the raise and lower classification on measurements within the DI. Reason: • Raise and Lower AGC regulation enablement prices differ markedly, with Lower much easier and less costly to provide.	• Separation of Raise and Lower response was used as an input assumption and not investigated in detail. • A combined approach to performance measurement would have benefits in terms of simplicity, transparency and ease of implementation. • While the difference in AGC reg prices between Raise and Lower is clear, experience may show that the distinction is not a compelling reason for separating Raise and Lower. We suggest a review of this and other issues in two years.

Element	Recommendation	Additional comment
Unit Aggregation	Settlement at the unit rather than portfolio level for both performance payments and AGC regulation cost recovery. Reason: • Portfolios should not be offered a settlement advantage over stand-alone units.	• This is the simplest, and fairest and most robust approach as compared to the current portfolio-based aggregation approach. • The proposed performance-based incentive arrangement is two-sided and therefore offers no specific advantage for portfolio-level settlement. • However, the proposed one-sided method of cost recovery for AGC regulation payments could advantage portfolio-level settlement if it were supported.
Residual Calculation	Top-down/energy balance approach preferred over the bottom-up approach Reasons: • Theoretically and practically robust by exploiting energy balance relationships. • Requires no adjustments to balance positive and negative cash flows. • Simple and transparent and supports simplified approaches to performance measurement.	• Top-down energy balance Offers clear benefits over current Causer Pays bottom-up approach in terms of simplicity and robustness • Reporting of different residual components is useful but it can be done outside of the settlement process.
Financial Weight	Regulation price (Raise and Lower): Reasons: • If Raise and Lower are to be separated for pricing purposes, regulation price or regulation costs are the natural candidates • Using regulation price separates the volume requirement for AGC regulation from the volume of the requirement for frequency control, which includes PFR as well as regulation.	• Modified from regulation cost of the Draft Rule to more easily support different volumes to cover both PFR and regulation • A slightly different financial weighting approach would be required if Raise and Lower were to be combined.

Element	Recommendation	Additional comment
Scaling	The maximum (or minimum) metered Gross dispatch Error within a DI, separate for Raise and Lower Reason: • Gross System Error is judged to reflect to total requirement for frequency control services.	• This recommendation is tentative, subject to more detailed review by AEMO. • There are other scaling options tied to measurable values that may have more desirable scaling properties at the micro level e.g. by offering payments more closely in proportion to need, as distinct from the relatively flat payments using Gross System Error.
Regulation Enablement Cost Distribution	The used component to be allocated to negative performance factors in a DI. The unused component to be allocated according to a formula based on previous measurements, to be determined by AEMO after consultation: Reason: • The recommendation is an AEMC policy decision which aims to avoid allocating costs to the residual directly because the residual may shrink in size or disappear entirely under evolving market arrangements. We agree that this is a reasonable approach.	• Our original preference was to allocate this to non-participating parties (the current residual). Units with negative performance factors are already allocated the costs of positive performance. The additional cost of assuring that response (i.e. the cost of the unused component) delivers a broad market benefit and should be allocated to those who are least likely to respond inappropriately. We recognise that allocating the whole cost to the residual may not be seen as equitable or supportive of risk management opportunities by participants. The AEMVC policy position is a reasonable compromise as it minimises the chance of bad behaviour.

Additional recommendations

Following on from the investigations and analysis undertaken for the AEMC, we provide the following additional recommendations:

1. The performance incentive arrangement should be phased in over a period of 12 months to allow participants and AEMO to adjust their systems and review outcomes progressively. The settlement amounts could be scaled by a temporary "operator constant" ranging from 0% to 100% as follows:

    - Quarter 1: 0% : "Dry Run" operation with no payments due
    - Quarter 2: 25%: Full settlement amount scaled by this factor
    - Quarter 3: 50%: Full settlement amount scaled by this factor
    - Quarter 4: 75%: Full settlement amount scaled by this factor
    - Ongoing: 100%: Full settlement amount with no further adjustment.

    The arrangement for recovering the cost of AGC regulation enablement would be implemented immediately as these costs are incurred from day one. Cost allocation would be broadly like the existing arrangement and so the case for a phased implementation is weaker.

2. AEMO should have the flexibility to include a component in the system performance metric for time error correction, subject to consultation. Ideally the time error correction component would be of a simple linear ex-ante form, analogous to the frequency metrics.

3. AEMO where necessary may and should use aggregated data from an earlier trading interval to calculate long run values wherever appropriate. These values would be an estimate of parameters immune from any later settlement adjustments that might occur. Policy wording should make clear that this should apply where appropriate, not just as a last resort. Such date could include default contribution factors as well as system-wide slow-moving values such the statistical measures of system performance.

4. Performance factors should be calculated, stored and used as means (averages of 4 second values) over the DI rather than summations (totals). This suggestion does not substantively affect the proposed settlement calculation but could simplify implementation by avoiding the need for interpolation to align with 4-second SCADA measurements and supporting metering resolution finer than the current AGC standard.

5. Asynchronous regions should be assessed and settled independently for all scheduled units. That would imply that interconnectors should be treated as pseudo-generators and pseudo-loads for the calculation of contribution factors, irrespective of any inter-regional frequency control arrangements that may be in place. This in turn would mean that interconnectors would attract settlement amounts. As actual payment to interconnectors could be seen as undesirable, their settlement amounts could be lumped into the residual.

6. Frequency Performance payments should apply following contingency events. Despite metering errors in the first few seconds of a contingency, performance payments should continue to be paid in this case, as such errors will be relatively small (except for the causer of the contingency over the period of the contingency and recovery from it).

7. AEMO should be permitted to support participants suppling locally measured data suitable for settlement in a form and in a manner agreeable to AEMO. AEMO may impose conditions such as periodic audit requirements.

8. The final rule should make clear that the deviations from a straight-line trajectory that help to control system frequency will be recognised as positive plant performance. This will help to clarify the expectations for plant performance with respect to energy market dispatch and the requirement for Mandatory PFR by scheduled and semi-scheduled plant.

9. IES recommend that AEMO's consultation on the frequency contribution factor procedure could further investigate the following issues:

   - the interaction of time error correction and the measurement of unit frequency performance, including whether or how to include time error correction using a smoothed offset adjustment independent of the AGC; and

   - alternative and refined scaling options, by examining their micro impacts under different conditions of high and low usage

10. Topics that could be the subject of a review after experience of, say, two years of operation include:

    - system performance and cost experience over the initial period;

    - whether it is appropriate for the contribution factors and payments to continue to be separated for Raise and Lower response or whether they could and should be combined into a single service; and

    - improved financial weighting and scaling strategies.

Table of Contents

Table of Figures

Table of Tables

Glossary

Term	Meaning
AEMC	Australian Energy Market Commission
AEMO	Australian Energy Market Operator
AGC	Automatic Generation Control – In this report, this refers to the central control device operated and maintained by AEMO to control regulation enabled units
DI	Dispatch Interval (5-minute)
FCAS	Frequency Control Ancillary Services
FDP	Frequency Deviation Price (usually in $/MWh)
FPP	Frequency Performance Payment (usually in $)
MPFR	Mandatory Primary Frequency Response
PFR	Primary Frequency Response
SFR	Secondary Frequency Response

1. Introduction

1.1. Background

Following years of declining system frequency performance in the East coast electricity system, in 2020 the AEMC published a final determination mandating the provision of primary frequency response from scheduled and semi-scheduled generators. The determination also recognised that mandatory arrangements are not a complete solution.

On 16th September 2021, the AEMC published a draft determination describing a new rule for enduring arrangements to support the control of power system frequency and to incentivise plant behaviour that reduces the overall cost of frequency control during normal operation. The Draft Rule seeks to improve on the current causer pays process in three main ways.

- by making the measurements and settlement concurrent
- by weighting the performance factors by DI to reflect different market conditions
- by providing a positive incentive for good performers

SCADA metering at 4-second intervals (8 seconds in Tasmania[1]) would support such an arrangement, as it does under the current causer pays regulation cost allocation system.

Stakeholder feedback has indicated that there was not enough detail or analysis of the proposed new arrangements given in the Drat Determination. The AEMC has acknowledged this feedback and has engaged IES to investigate options for the incentives framework before coming to a decision.

1.2. Aim of this project

The aim of the project is to assist the AEMC investigate a range of options for a performance-based incentive arrangement to support the control of power system frequency. The results of the IES analysis will inform the AEMC and its stakeholders and support the selection of a preferred arrangement. This report presents details the process undertaken and the results of this work. The broad process to be followed is as follows.

1. Initial screening

The first step is to conduct a qualitative assessment to eliminate options that are inconsistent with the assessment framework. We perform a similar analysis on combinations of options to develop an initial set of candidate combinations.

2. Short sample period assessment

In this step quantitative analysis is conducted on the initial set by testing it over a representative two weeks of 4-second data from the full sample period. A final set of the best-performing combinations (candidates) is then selected for more detailed analysis.

[1] https://www.aemo.com.au/-/media/Files/Electricity/NEM/Security_and_Reliability/Ancillary_Services/Regulation-FCAS-Contribution-Factors-Procedure.pdf

3. Long sample period assessment

In this step, the final candidate combinations are tested using data from the entire sample period of 11 months. It is envisaged that a maximum of four combinations (scenarios) will reach this stage.

The analysis will use historical SCADA and market data from 1 February 2021 to 26 December 2021 (Inclusive) which will give an initial indication of the effect of the proposed arrangements.

This is a technical report. It assumes that the reader is familiar with the concept of frequency control and the AEMC and AEMO consideration of this topic in recent years. The AEMC Draft Determination and Rule is the key reference.

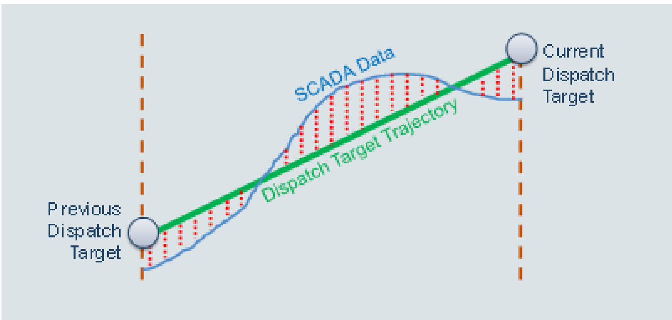
1.3. Overview of proposed incentive arrangements

1.3.1. Outline of the proposed incentive arrangements

While meeting the goal of good frequency control, there are many design choices or an incentive arrangement. This section outlines the broad approach to the proposed arrangement, updated from the Draft Rule.

The diagram below shows a straight-line scheduled energy market trajectory of a dispatchable unit, together with its SCADA-metered actual output. This trajectory is one of several possible choice for a reference trajectory.

Figure 1: Example of Reference Trajectory and Outcome



$$\mathit{Scheduled\ Deviation\ Gen} = \mathit{ActualGeneration} - \mathit{TargetTrajectory}$$

Source: AEMO Causer Pays Procedure

The deviation of the unit in MW is shown as the red-dotted shaded area. One possible incentive arrangement would sum a product of some metric based on, say, frequency or dispatch error measurements, with measured MW deviation. We can consider this as being done over a whole billing period, although in practice we will break this up into DI measures for weighting purposes and for DI-level settlement.

The reference trajectory can be regarded as the contracted MW quantity of a very short-term swap contract that is automatically issued to scheduled units at the start of a DI. Subject to any other rules that may apply, a unit may deviate above or below that level and be rewarded or charged on top of any energy market adjustments, depending on the extent to which that deviation supports frequency control. Clearly, the reference trajectory and metric are key components in any performance incentive arrangement.

1.3.2. Proposed settlement formulae

The basic form of a DSCP/FDP/frequency performance payment (FPP) function for a single DI can be represented as:

$$FPP = (contribution\ factor) \times (financial\ weighting) \times (scaling\ factor) \qquad (1)$$

- The *contribution factor* is a value that represents the contribution (or impact) of a metered unit to the system state. It is determined based on a function of a system performance metric and a unit contribution metric. It can be represented mathematically as

$$contribution\ factor = f(system\ metric, unit\ metric) \qquad (2)$$

- The *financial weighting* is intended *to* price performance in a way that maintains interest in performing frequency control relative to operating in the energy market and to ensure long term investment. It could be based on:
    - market measures such as enablement price or energy price; or
    - cost-based measures
- A *scaling factor* may be determined automatically based on system metrics or specified manually.
    1. An automatic *scaling factor* would be derived from system metrics such as the need for (or usage of) regulation services during a DI period (draft rule)
    2. A manual *scaling factor* may be specified by AEMO, or otherwise, to achieve a system objective.

Reviewing the proposed settlement formula:

- *Contribution factor* will likely be some form of ratio; it represents a “share” and has no units.
- *Financial weight* will have units of $/MW or can be converted to $/MW by multiplying by the duration of the DI; so that
- *Scaling factor* will be in MW (perhaps multiplied by a dimensionless factor) to give an FPP in $.

1.4. Outline of this report

The remainder of this report is structured as follows:

- In Section 2 we review the methodology used during the project. The various options under each procedure element are set out, along with the criteria used to assess them
- In Section 3 we describe the process the project team went through to arrive at a preferred approach to the most critical design elements. A few critical elements - the performance metric and the closely related issue of scaling the settlement payment - occupied most of the time and energy of the project team. Further detail on these critical elements and a summary of the approach to the less difficult elements have been assembled in APPENDIX B.
- Section 4 outlines the scenario selection process and the elements of the scenarios that were chosen to run at various stages. While this is reported separately for clarity, in practice these scenarios were developed as an integral part of the process described in Section3.
- In Section 5 we present the results of short period performance factor and settlement runs which used data spanning two weeks from September 2021, with commentary and conclusions. These results guided the development of a series of modified options for critical procedure elements.
- Section 6 outlines the results from a final set of three scenarios which delivered credible turnover volumes. Each of the scenarios had some different elements within that requirement and they warrant further consideration at the implementation stage.
- Conclusions and Recommendations are in the Executive Summary.
- Appendices provide background analysis and results
  - Appendix A describes the elements of the settlement calculation procedure
  - Appendix B provides more detail on the initial investigations that were carried out, focussing on the outcome rather than the process outlined in Section 3.
  - Appendix C sets out a way to achieve gross system error levels of scaling under ad FDP arrangement, highlighting the relationship to the basic settlement formula of the Draft Rule, as modified by this project.
  - Appendix C describes a correlation analysis on a proposed PFR metric
  - Appendix E: contains the short period result charts discussed in Section 5
  - Appendix F: contains the long period result charts discussed in Section 6.

2. Methodology

2.1. General Approach

Due to the potentially large number of combinations that could make up the scenarios under investigation, and the incoherency of combining some elements, a staged approach to the assessment is required. At each stage some options or complete scenarios would be removed from further consideration after an appropriate level of analysis. The three stages summarised in Section 1.2 above are outlined in more detail below.

2.1.1. Initial Screening

At this stage, the options (or group of options) were assessed independently of all other options/elements. The procedure is not synthesised, nor are any revenue/cost calculations made. The purpose of this stage is to remove clearly unworkable options from further consideration to reduce the number of scenarios for later investigation.

Although this stage is labelled “Initial”, the investigation of the procedure was iterative: the project team would synthesise new options or scenarios based on results or investigations of other scenarios. This meant, the team would often return to a principles-based assessment when new scenarios or options were synthesised.

2.1.2. Short Sample-Period Assessment

After the first stage, an initial set of scenarios (candidate procedures) was determined. These scenarios were implemented in the calculation engine and settlement results and unit contribution factors were produced over a relatively short period. This work was carried out in the following phases:

- The performance metric was investigated by
    - Charting and reviewing the 4-second values (for the proposed metrics) in NEOPoint
    - Performing correlation calculations of the metric with other measures for system need and/or system performance (e.g., regulation signal, net deviation of enabled units etc)
    - Implementing base scenarios into the calculation engine. These were developed from performance metrics identified as more favourable over others, i.e., only the performance metric differed between each scenario. The contribution factor results (spanning 2 weeks from 1 Sep 2021) were calculated and investigated.
    - Discussing and obtaining feedback from stakeholders through the TWG process
- Once the preferred metric was identified, variations on the preferred base scenario were implemented into the calculation engine and investigated.

- Settlement results were calculated for a short period spanning 2 weeks from 6 Sep 2021[2]
- Through this stage, options for most of the elements of the procedure not determined through the initial screening were investigated and determined. Options for scaling were not determined.
- Scaling was identified as a major part of the procedure and the next phase of the investigation was devoted to Scaling alone.

2.1.3. Long Sample-Period Assessment

After further review, the project team settled on three scenarios for a full 11 months of settlement calculations. The period covered was 1 Feb 2021 to 26 Dec 2021 (Inclusive), equivalent to 47 settlement weeks.

This long period analysis was expected to highlight any conditions or issues that were not evident in the shorter analysis.

2.1.4. Commentary of the effectiveness of the approach

The strategy behind this project had to evolve as work progressed. Items that required more than expected analysis were the choice of an appropriate metric and the way that any relative performance measure should be scaled to support actual payments. This will be evident in the remainder of this report.

As a result of the work put into what appeared to be the key topics, some options for other elements were not considered in any scenarios.

There are significant uncertainties around key elements of the arrangement that cannot be fully resolved using historical data or in practice by any other method than experience. The most important is the potential response to participants to the incentive.

2.2. Scenarios

2.2.1. Introduction

A scenario is a combination of options that together constitute a viable system to incentivise good frequency control. The steps required to evaluate a scenario are:

1 Specify and calculate the frequency performance metric (system metric)

2 Specify and calculate active power deviations for metered units (unit metrics)

3 Calculate the impact of non-metered generation and load (residual)

4 Calculate contribution factors

[2] The 6th was chosen as it was a Monday; All settlement results span entire weeks and begin on a Monday. The reason is that the options that require long-term values (e.g., long-term gross error or default contribution factors) used the average of the relevant measures from within the current week (starting on Monday) as a proxy.

5 Specify the treatment of enabled units, including the nature of their participation in the arrangements and how enablement costs will be paid for

6 Calculate settlement amounts

7 Describe and evaluate the effectiveness of the process scenario

APPENDIX A contains a process diagram for these steps. The options for each element identified at the planning stage of the project are outlined in the following sub-section. To keep the description of a scenario manageable, we also devised and describe a classification system. However, as the project progressed, an initial assessment allowed us to drop or relatively easily settle on many options. Some of the classification and naming has changed since project inception as new options were synthesised through the investigation process. Most of the effort in the project was in resolving the approach to two elements: choice of the performance metric and choice of the settlement scaling method. Options not identified at the project start were also identified and analysed.

2.2.2. Elements, Options and Parameters

A candidate procedure for calculating frequency performance payments can be decomposed into multiple **elements**, each element is associated with a list of **options** which could be chosen for the final procedure. Each option may have associated **parameters** whose magnitudes are set by the operator within a defined framework or determined as part of applying the procedure. For example, an element of the procedure could be the choice of performance metric which could be one of a few options such as FI, RR, smoothed Hz, etc. The smoothed Hz performance metric will have as a parameter the time constant of the smoothing function. The elements of the procedure are listed with their corresponding identifiers in Table 2-1 (See APPENDIX A for detailed discussion on the elements).

Table 2-1: Procedure elements and identifiers

Name	Identifier
Performance Metric	PM
Reference Trajectory	RT
Classification	C
Unit Aggregation	UA
Residual Calculation	RC
Financial Weight	F
Scaling	S
Regulation Enablement Cost Distribution	EC

2.2.3. Options by Element

The options for each element identified are listed in the following tables. The options are adapted from the options provided by the AEMC. Options for the performance measure, reference trajectory and the classification element are provided as a list of base options and modifiers. For example, PM3.1.0 refers to the performance metric calculated as “$Sign(FI)$”

(with no deadband) while PM3.0.1 refers to the performance metric calculated using FI with a dead band. Some combinations are easily ruled out.

The tables following contain the initial set of options considered and are provided to show the starting point for the investigations. As the work progressed, some additional options were developed and analysed and some were dropped quickly on first principles grounds. The process we followed for this is outlined in Section 3 and detail provided in Appendix B.

Performance metric

Table 2-2: Performance metric base options and modifiers

Base Metric	Performance function modifier
PFR only – Instantaneous Frequency (**PM1.X**)	Product function (**PMX.0**)
SFR Only – Smoothed Frequency (**PM2.X**)	Sign-product function (**PMX.1**)
Hybrid Factor – FI (**PM3.X**)	Sign-deadband function (**PMX.2**)
Separate Factor – PFR, SFR applied separately (**PM4.X**)	Correlation function (**PMX.3**)
ACE – as calculated by AGC (**PM5.X**)	
Gross Dispatch error (**PM6.X**)	
Net Dispatch error (**PM7.X**)	
PFR adjusted by time-error offset (**PM8.X**)	
SFR adjusted by time-error offset (**PM9.X**)	

Reference trajectory

Table 2-3: Reference trajectory base options and modifiers

Base Trajectory	Droop Modifier	Regulation Modifier
Target-target linear trajectory (**RT1.X.X**)	No Droop (**RTX.0.X**)	No AGC-REG signal (**RTX.X.0**)
Initial-target linear trajectory (**RT2.X.X**)	With droop (**RTX.1.X**)	With AGC-REG signal (**RTX.X.1**)

The AGC basepoint signal was also considered as a reference trajectory but was not implemented in the calculation engine. AEMO investigated this option and determined that it inappropriate as a trajectory.

Classification

Table 2-4: Classification base options and modifiers

Base Classification	Direction Modifier	Enablement Modifier
Positive or Negative classification before 5min aggregation (**C1.X.X**)	No Raise or Lower classification (**CX.0.X**)	No enablement classification (**CX.X.0**)
Positive or Negative classification after 5min aggregation (**C2.X.X**)	Raise or Lower classification before 5min aggregation (**CX.1.X**)	Enablement classification after 5min aggregation (**CX.X.1**)

Base Classification	Direction Modifier	Enablement Modifier
	Raise or Lower classification after 5min aggregation (**CX.2.X**)	

Other Elements

Table 2-5: Options for all other elements of the procedure

Unit Aggregation	Residual Calculation	Financial Weight[3]	Scaling	Enablement cost distribution
By DUID (**UA1**)	Bottom-up (**RC1**)	Separate regulation (**F1**)	System Need (**S1**)	Used from poor performers and unused from all (**EC1**)
By Portfolio (**UA2**)	Energy balance (**RC2**)	Combined regulation (**F2**)	Responsive fraction (**S2**)	All from poor performers (**EC2**)
		Energy price (**F3**)	Operator's Constant (**S3**)	All from non-metered (**EC3**)
			Gross Error over enablement (**S4**)	Used from poor performers and unused from non-metered (**EC4**)
			Gross Error over long-term gross error (**S5**)	Used cost from poor performers, unused cost accorded to some average of prior performance[4] (**EC5**)

Parameters

The parameters for the above options are listed below

Table 2-6: Identified parameter information

Name	Possible values	Relevant options (or modifier)
SFR Frequency lag	35sec	**PM2**, **PM4**
Relative weight of SFR v PFR	1:1	**PM4**
Frequency bias	2800[5]	**PM5**
Dead band	15mHz	**PMX.X.1**
Droop setting	AEMO/DUID agreed value up to 5%	**RTX.1**
Operator's Constant		**S3**

[3] With suitable floors/modifiers such that the financial weight is appropriate. For example, it does not make sense to weight the FPP payments with a negative value.
[4] The long-term factors are also used as default contribution factors if the 5-min factors are unavailable for any reason
[5] From Regulation Causer Pays Procedure

2.2.4. Scenarios

A **scenario** under investigation is a unique combination of options (1 per element of the procedure). Different values for parameters will also be investigated but these can be considered either as separate scenarios or under a sensitivity analysis of the same scenario. For example, the procedure represented by {PM3.1.0 (Sign of FI), RT1.0.0 (target-target only), UA1, RC2, F2, S3, EC3} is a single scenario under the investigation.

2.3. Assessment Framework

The assessment framework is expected to be adjusted throughout the life of the project. However, the project will be guided by the principles determined by the AEMC in its draft determination to the PFR incentives rule change, namely:

- Promote power system security and reliability
- Appropriate risk allocation
- Technology neutral
- Flexibility
- Transparent, predictable, and simple
- Minimise implementation costs

The AEMC provided additional guidance in its Request for Proposal for this Project. Operational objectives for the incentive arrangements are to encourage:

- Frequency responsiveness - to deliver a large proportion of plant that is responsive to frequency in accordance with agreed settings à better primary control
- Dispatch compliance - reduced dispatch deviations which drive the need for regulation response (measure of systemic error)
- Regulation performance - Do we expect the behaviour of non-enabled units to correlate with AGC reg? (This may create transparency and operational issues)

3. Initial Investigations and Issue Resolution

3.1. Introduction

The project team for this project consisted of staff from the consultant, IES, with advice and input from AEMO and AEMC staff members as well as, critically, comment from members of the frequency Control Technical Working Group over several meetings held as the project progressed. While the brief for the project was clear and reaching a conclusion on most options relatively straightforward, choice of the *performance metric* and *scaling of the settlement formula* were two related issues that required a series of iterations between "principles" analysis and scenario runs. This section describes that process and the thinking behind it in a way that is less evident in the option level summaries provided in Appendix B.

This section is of course the consultant's interpretation of the process.

3.2. Initial Perspectives

As noted in the introduction, the motivation for this project was to provide more detailed analysis of how the proposed incentive arrangement outlined by the AEMC draft rule would work in practice. We can discern two broad perspectives on Draft Rule:

- More detail is sought on the short-term operational incentives the arrangement would offer. The short-term perspective is concerned about devising incentives that drive proportional and stable responses sufficient to maintain the frequency standard, noting that the arrangement will be an overlay on the mandatory PFR (MPFR) rule as well as the long-standing AGC regulation arrangement.
- There is a desire to ensure that the turnover that would result is sufficient to promote the required long-term investment in frequency control. The long-term perspective focusses on the dollar turnover of the arrangement, including its relationship to the existing AGC regulation turnover.

These perspectives should not be mutually exclusive, but they also may not automatically align.

While ensuring adequate turnover was a common theme, it was also canvased in some detail in AEMO's discussion paper prepared for the AEMC rule change process[6]. That paper hypothesised that, in a system with tight dead band PFR, frequency deviations are small and possibly not a good indicator of the work units are doing to control frequency. Specifically, in the AEMO submission on the AEMC Draft Determination, AEMO argues:

[6] AEMO 2021 - Primary Frequency Response Incentive arrangements - Discussion Paper

> Where control of frequency is performed by ubiquitous tight deadband MPFR, the Discussion Paper noted that this has the effect, depending on the performance of AGC-REG, of better aligning dispatch errors to one another, because they can be corrected without noticeable, accumulating deterioration in frequency and accumulation of time error. The implication was that frequency may not be a good measure upon which to credit and debit deviations, and instead deviations could be priced irrespective of frequency.
>
> For these reasons, the recommendations for changes to 3.15.6A in the Discussion Paper were premised on using data to assess dispatch errors (or imbalances) and crediting or debiting these. The credits and debits are indexed to, and scaled against, the prevailing cost of regulation FCAS, a service designed to correct dispatch errors (resolve imbalances). These recommendations were based on the assumption of a high ratio of proportional active power response to changes in power system frequency (MW/Hz) provided by a large proportion of generation plant, not assuming scarcity of primary response.

and:

> The AEMO Discussion Paper recommended scaling the costs (although not using RR) because dispatch errors may exceed EA. Without offsetting errors, MPFR will act to slow the deterioration of frequency. The primary response will respond in proportion to the change in frequency and although secondary control will eventually need to correct any persisting error, primary response could exceed the EA. The purpose of the scaling factor is to allow performance payments to exceed the cost of Regulation FCAS and was thought consistent with introducing payments for mandatory primary response, which could exceed EA.

A largely short-term perspective on a frequency control incentive arrangement is in the reports of a 2021 research project into a Double-Sided Causer Pays arrangement carried out by IES, sponsored by the AEC and two market generators and supported by ARENA[7]. This project focused on the design of a short-term incentive arrangement based on frequency deviation pricing (FDP) principles. While sponsored by the AEC, the report does not necessarily reflect the views of the project sponsors; their support for the project was intended to inform debate.

While the IES proposed approach was not inconsistent with the needs of the long term, it took the view that the settlement formula could simply be scaled up to deliver the required turnover. This could be easily done because the arrangement was inherently balanced physically as well as financially, regardless of the scaling. The project also took as a given that a direct control of frequency and time error was the objective and that, after review of SCADA metering issues, the values required for settlement could be acceptably measured. However, this perspective is incomplete because in implementation and operation AEMO cannot be given total freedom to scale settlement payments; scaling must be set by some measurable and defendable process.

In summary, these short-term and long-term perspectives led to divergent *initial* views on the nature of a good performance metric. Further, the short-term operational approach based on

[7] https://www.energycouncil.com.au/media/wzpjzrjs/dscp-overview.pdf

frequency deviations could not be sustained without a robust scaling mechanism. These matters dominated the work of the project team, as described in the following sub-section.

3.3. Process for Resolution

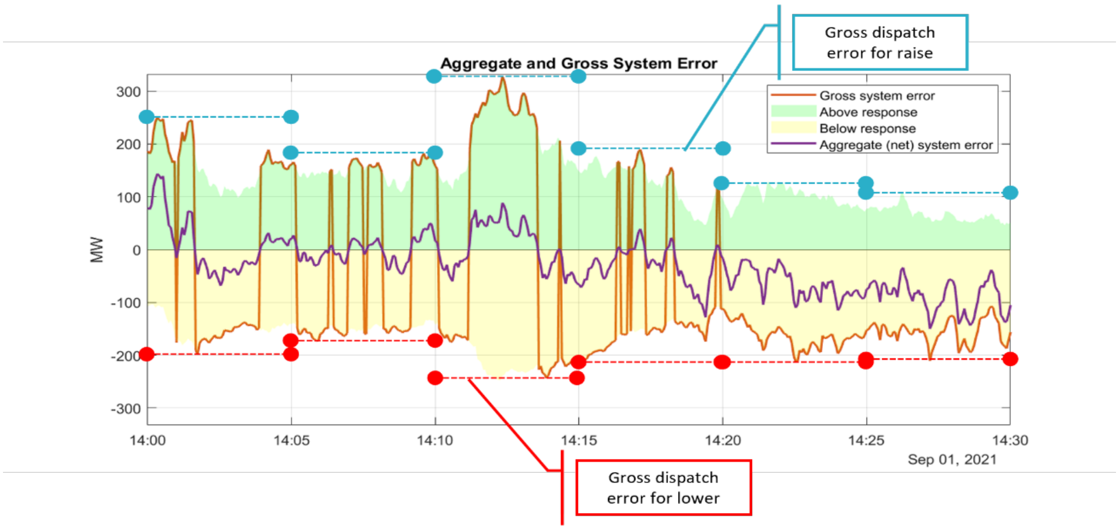
3.3.1. Selecting candidate metrics for settlement analysis

Th first step in the process was to consider an arrangement paying and charging for PFR, including Mandatory PFR. We considered a range of options based on frequency beginning with a correlation measure, which could be used to pay according to its MPFR obligation once a threshold crossed. However, we found from analysis (see APPENDIX D) that a significant correlation between frequency deviation and unit MW deviation at the DI minute level could not be found due to the variability of both measures within the DI and the small sample size (75 on the mainland). Significance improved at the half hour level and long, but a basic policy starting is to settle based on measured within each DI to avoid potential conflicts in unusual situations.

Another frequency-based measure was to aggregate the product of the raw frequency, which at 4-second intervals approximates the response time expected from MPFR. We found that this measure was highly volatile at the DI level, although it improved at half an hour or longer so that, over a billing period of a week the measure, when averaged over all samples, was an excellent approximation to the statistical measure of covariance. This could be a workable measure for payment as it captures both scale and quality of performance. Again, it is only robust when assessed over a long period, so potentially problematic on that score.

At the same time, we began exploring the AEMO concept of using dispatch error measures to assess performance and the scale of payment rather than frequency. Examination of 4 second raw and processed data using NEOPoInt was helpful in this work as we could view and share different time periods and different durations easily. The chart below illustrates the essence of this work.

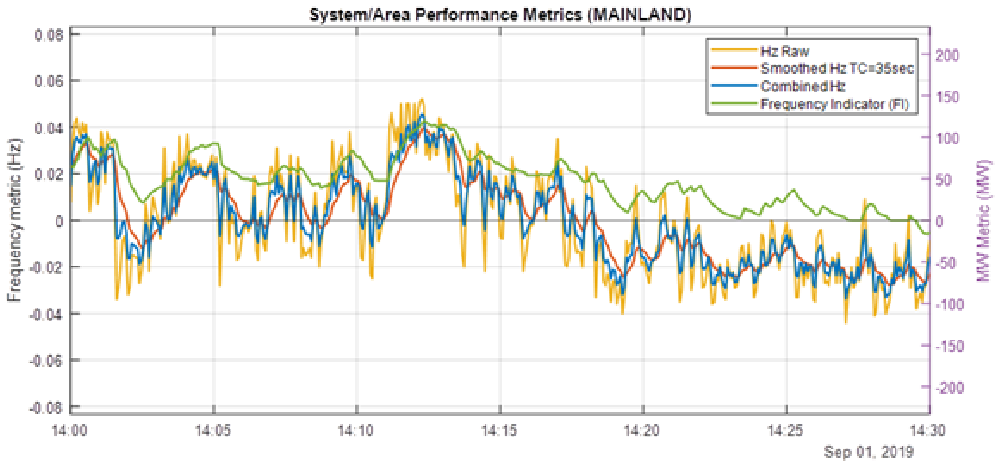
Figure 2: Gross system/dispatch error scaling



The shaded areas of the chart shows the sum of all the positive (green) and negative (yellow) unit responses over the period. The difference is the purple net system error line. This measure is assumed correlated with the frequency error and so an approximation to the scale of the responses drive by the AGC regulation system. Assuming that all net (or residual) deviations are bad and a cause of deviations, we could hypothesize that the "majority" are working to correct frequency deviations. Without delving into the details here, this delivers the orange line as a potential metric, much larger than the "net" drive metrics (presumed to track the purple line reasonably closely) based on frequency. Also shown are the DI raise and lower values if this measure were to be used for scaling in the settlement formula. Note that both PFR and secondary response by AGC-enabled units and others are all boiled into this metric.

We also gave considered a suitable frequency-based metric to support a sustained unit response to disturbances, even when the raw frequency deviation has returned to near zero. In a PID controller this would be the integral component. Based on the work of the AEC project, we proposed a smoothed version of frequency as a possible metric, somewhat like the AGC regulation Frequency Indicator (FI) measure but simpler and more easily measured. Further, given that fast and sustained responses are both important, a combined Hz measure (the mean of a raw and smoothed metric component) was also identified for further analysis. A sample of these metrics is shown in the figure below, covering the same period as the previous chart. Note the roughly 100 MW range each side of zero in the case below, and the range of the gross system error metric which is about three times larger.

Figure 3: A sample of frequency-based metrics



In summary, this analysis identified four base scenarios built around the following four metrics:

1. Frequency Indicator (FI), not as a realistic option but as a reference case
2. Smoothed Hz, to reward and encourage sustained (secondary) response
3. Combined Hz, to remunerate PFR as well as secondary response
4. Gross system error, as outline above

Our initial set of runs focussed on performance measures only – not settlement. Further, to allow comparison independent of any scaling issues we normalised the results, which are reported in Section 5. On review of the summary result as well as at a more detailed level, we concluded that the different performance metrics gave outcomes that were broadly similar, although not in every detail. This cleared the way for doing settlement runs with these metrics

3.3.2. Settlement analysis and the focus on scaling

The proposed AEMC settlement formula normalises the performance measures to deliver a fixed contribution factor (CF). These factors would be used to allocate a dollar amount determined by

- reference price (chosen as the regulation price for the DI, expressed in $/MW per hour or $/MWH);
- a scaling factor which is most simply be expressed in MW
- The DI duration expressed in hours.

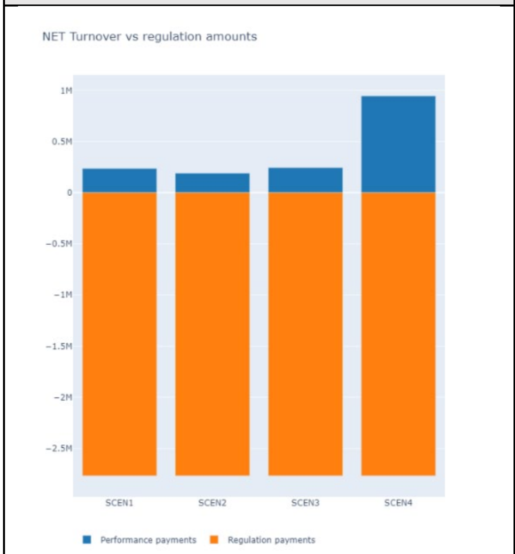
See section 4 for details. In the first set of settlement runs spanning two billing weeks in the first half of September 2021, the scaling factors were:

- for the first three Scenarios the maximum in the case of Raise (or minimum in the case of Lower) of net system error over each DI.
- For the fourth scenario, the scaling was gross system which , as noted above, is typically much larger than the net value.

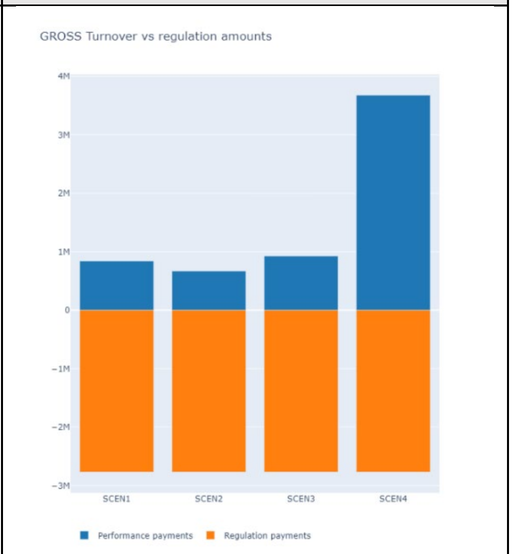
The chart below shows net and gross turnover. "Gross" is essentially positive payments in each DI added up over the period while "net" nets out and negative DI payments by a unit over the period. This reflects the unit's bill.

Figure 4: Net and gross turnover for performance an AGC regulation for base scenarios

Net turnover in S



Gross turnover in $

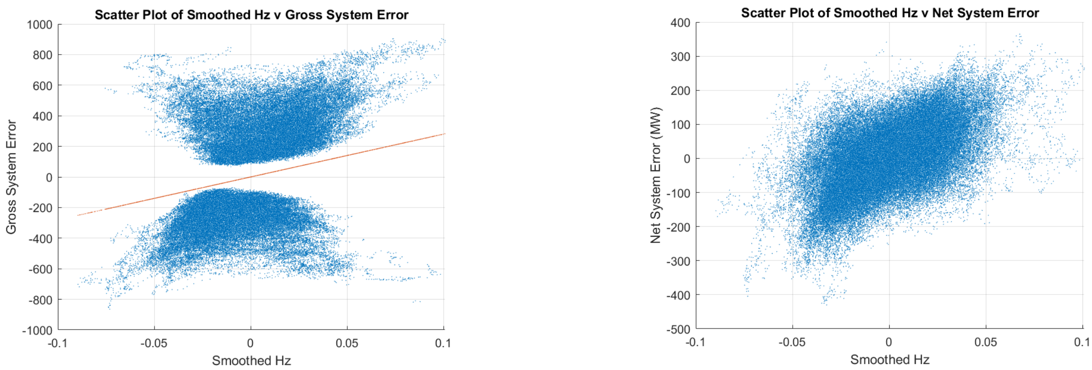


The results show very small turnovers for the Scenarios 1-3 using “net” scaling but much larger turnover for Scenario 4 using “gross” scaling, in both the gross and net reporting. This was no surprise as scaling was the only differentiating factor. If the aim is to have a turnover at the DI level of the same order as AGC enabled regulation, gross scaling clearly does a better job.

The real surprise was the degree of shrinkage between gross and net reporting, reducing net turnover to about a third of gross for all scenarios. This implies a lot of negative performance, even for relatively well performing units.

To explore this issue further we constructed some scatter plots of each metric compared with smoothed frequency. These are shown in the figure below. The gross error plot is on the left and the net on the right. Both are compared with smoothed Hz on the horizontal axis. Note that the MW scale on the gross chart is much larger than on the net chart.

Figure 5: Scatter plots of smoothed frequency v gross and net system errors



We knew from earlier work on such as the example in Figure 2 that gross system error tended to have binary behaviour, oscillating from large positive to large negative values with not much in between. The plot on the left confirms that. The modest surprise in this plot is how poor the relationship is between gross system error and a reasonable expectation of requirement as measured by smoothed MW.

The net system error case on the left shows a clearer relationship but the scatter is wide. Yet the right-hand plot presents the current system behaviour and the system functions.

We looked at explanations for this poor correlation including the time error frequency offset correction of the AGC regulation system. There were observable differences but the same picture emerged.

To investigate further we constructed a simple simulation where some fraction of the units were performing PFR only (for simplicity). However, they were also programmed to suffer from noise; that is, they did not respond to frequency perfectly but tended to wander away and then return. Plotted against raw frequency, we were able to reproduce the general shape of the charts above.

This work confirms that there is much random noise in the system which creates a band of MW noise on the positive and negative sides which affect and expands the size of the gross system error. However, this noise component is largely washed out in two directions:

- within a DI, by noting that the calculated payments for a DI will total a net turnover based on the aggregate (net) error.
- For each unit over time, by noting that its positive payment DIs will be offset by its negative payment ones. It is this at the DI level that drives the shrinkage we observe when going from gross to net.

From this block of work we could reasonably conclude the following:

- The use of gross system error as metric appears to be unsatisfactory because of its poor alignment with underlying objective, which is frequency control. A different forms of gross system error using gross system error as a metric was also considered but rejected because of its poor incentive characteristics.
- We concluded that a frequency-based metric is the appropriate metric to use. The current relatively small deviations do not appear to be a roadblock as we could produce performance assessments that made sense. Using SCADA metering is a challenge but workable if some allowance is made for special cases.
- Given that the team considered the combined metric to be the best of the proposed frequency metrics as it dealt with both PFR and regulation type responses.

Finally, and importantly, gross system error was still considered to be an appropriate metric. This was so despite the fact that the measure contains a great deal of random noise, at least at present. The incentive procedure must live with random dispatch error because:

- it is a fact of life at present;
- in any case, it is impossible to distinguish intended or poor performance deviations from random ones over a short interval such as a DI;
- randomness tends to net out settlement amounts over a full billing period; and
- subject to further study, the new incentive mechanism, improvements to AGC regulation and the continued roll-out of MPFR across units, including inverter-based units, should lead to any perceived overpayment from using gross system error being self-correcting.

There is further work to do in refining the scaling factor to ensure it is not needlessly volatile on a per interval basis, yet still reflect scale of response on the system that needs to be paid for. An option worth investigating is calculating the gross dispatch error, not with reference to the net dispatch error (i.e. on the side of the majority), but by the chosen performance metric. This was not possible in this project due to performance measures and scaling factors being calculated independently.

3.4. Conclusions on initial investigations

With the preferred performance metric and general scaling approach resolved, the remaining scenario runs and additional analysis could focus on refining these performance metric and scaling options (but mostly scaling), together with some variations on other design elements. Even then, we recommend that further analysis of scaling at a micro level should be pursued at the time of implementation.

Scenarios run and conclusions reached are presented in Sections 5 and 6, with a full set of charts in Appendices F and G. APPENDIX B contains a more compete discussion of the individual options for each procedure element, with analysis where required. For ease of reference, the table below summarises the preferred options for each element, noting where more work may be required at the time of implementation.

For completeness, the following table sets out the options chosen for the final runs. The rationale for most of them can be found in Appendix B.

Table 3-1: Element options for used for long period Scenario runs

Element	Option
Performance Metric	Combined Hz
Reference Trajectory	Target to target; no AGC reg except for Scen. 3.15
Classification	Separate Raise and Lower; performance classification made at measurement level
Unit Aggregation	Unit level settlement
Residual Calculation	Top down / Energy balance
Financial Weight	Regulation price
Scaling	Several options tried with two preferred. More analysis required.
Regulation Enablement Cost Distribution	Used to negative factors, unused to negative factors average over an historic period.

4. Scenario Selection Process

4.1. Overview

Based on the filtering of scenario elements performed as part of the initial investigations, a set of complete scenarios and variations were selected for further examination. Note that scenario selection was an iterative process, with one set of analyses informing the next, e.g. the conclusions from the analysis of the base scenarios informed the selection of the scenario variations and so on.

This section provides a description of the selected scenarios, along with brief conclusions from the assessments.

4.2. Base Scenarios

Table 4-1 specifies the four base scenarios, whose main variation is the performance metric. After initial investigations of their behaviour and practicality, the following performance metrics were selected for further assessment:

1. **Frequency indicator (FI):** the performance metric used in the current causer pays approach and selected as a baseline to facilitate comparisons.
2. **Smoothed Hz:** a single metric that exhibited a reasonably high correlation with both raw Hz (reflecting PFR) and FI (reflecting AGC regulation).
3. **Combined Hz (weighted average of Raw Hz and smoothed Hz):** a blend of fast and slow frequency metrics that tries to capture both PFR and regulation requirements.
4. **Gross dispatch error:** a metric based on system wide gross deviations selected as a potential alternative to frequency-derived metrics.

Note that the regulation enablement amount (EA), the denominator of the scaling factor in the table, is meant to cancel out the EA in the Draft Rule TSFCAS and to replace it with another MW amount. This was later simplified by a decision to use Regulation price rather than regulation cost as a financial weight. Under that convention, the denominator would simply be removed.

The base scenarios were investigated in the following short sample period simulations:

- Performance factor analysis
- Short period settlement analysis with financial weighting and scaling as envisaged in the Draft Rule

The performance factor analysis (refer to Section 5.2) showed a broad similarity in the allocation of performance factors between different types of plant across all performance metrics. However, Scenario 3 (combined Hz) was selected as the preferred base scenario for the following reasons:

- a frequency-based metric is more transparent and consistent with the AEMC's preference in the Draft Rule for the performance metric to be a measure of the need to raise or lower the frequency of the power system;
- it can easily be measured and tracked locally by participants; and
- the combined Hz measure reasonably captures both the fast and slow components of frequency behaviour.

The short period settlement analysis (refer to Section 5.4) indicated that the financial weighting and scaling as described in the Draft Rule would lead to low net frequency performance payments (relative to regulation enablement payments). Different settlement formulae and scaling factors would be explored in the subsequent analyses.

4.3. Variations on Preferred Base Scenario

Table 4-2 specifies four variations on Scenario 3, the preferred base scenario. These scenarios consider variations on other parameters including the reference trajectory, inclusion of time error correction frequency offsets and alternative settlement formulae and scaling factors.

The results of the short period analysis (refer to Section 5.4) were as follows:

- Scenario 3.1 (including time error Hz offset): as expected, non-enabled plant exhibited a poorer response as control to local frequency (without time error correction) is not always aligned with frequency + Hz offset. This was not considered further given the AEMC's preference for a transparent and simple performance metric.
- Scenario 3.2 (including regulation component in reference trajectory): the results were largely inconclusive as the default scaling factor leads to small frequency performance settlement volumes.
- Scenario 3.3 (FDP – IES-AEC approach): showed that settlement volumes could be increased significantly with appropriate scaling. However, the scaling approach in this scenario was an operator constant that would be difficult to justify.
- Scenario 3.4 (relative performance): also exhibited higher settlement volumes due to scaling, but the scaling approach was again difficult to justify.

The scenario variations showed that the choice of a robust and justifiable scaling factor is critical and warranted further investigations in the final set of analyses.

4.4. Scaling Factor Variations

Table 4-3 specifies three variations on scaling factor approaches:

- Scenario 3.8: Gross system error scaling – scaling reflecting the need to cover the total volume of helpful active power deviations across the power system.

- Scenario 3.14: Hz spread scaling – an extension of the FDP approach where the operator's constant is calculated based on the value of an active power deviation at a reference frequency (refer to Appendix B for more details).
- Scenario 3.15: Inclusion of AGC regulation trajectory in the reference trajectory. Because AGC units are largely excluded from the incentive arrangements under this approach, the requirement based on gross system error is reduced by the metered AGC regulation component of performance. This reflects the work potentially required from the non-enabled units.

In this table:

- *Gross error* is the sum of all positive (for raise) or negative (for lower) deviations from the reference trajectories for all units. This represents the total amount of response in the system.
- *Net error* is the sum of all deviations (positive or negative) from the reference trajectories of all scheduled units.
- *freqref* is a reference frequency (standard deviation on mandated maximum dead band); we have chosen the mandated maximum dead band of 15mHz for the current simulations
- *AGG* is the sum over all scheduled units with positive performance factors (the gross sum)
- *long-term* is a value calculated in advance over a sufficiently long period, and certainly more than one billing period, and made known in advance
- The ex-ante and ex post column describes whether the metric can be tracked in real time (ex-ante) or not (ex post)

The long sample period analysis (refer to Section 6) concluded that all of the scaling factors considered produced reasonable settlement amounts. The scaling factor based on gross system error (Scenario 3.8) was selected as the preferred choice for the following reasons:

- gross system error is an easily calculated and transparent value that is based on the measurement of all active power deviations for metered plant; and
- it is independent of AGC and will not be affected by changes to the tuning of AGC.

Scenario 3.15 was not preferred based on separate considerations as discussed in APPENDIX B. However, Scenario 3.14 warrants further investigation and comparison with 3.8 at a detailed level, for example under situations of high and low demand for frequency control. Outcomes in these extreme but important cases could be different, providing different incentives and therefore relevant to a final choice.

Table 4-1: Procedure elements and identifiers for Base Scenarios

ID	Performance metric	Contribution factor4F[8]	Reference trajectory	Financial weighting	Scaling	Commentary
1	Frequency indicator (FI)	$\frac{\textit{Unit sumproduct}}{\textit{gross sumproduct}}$	Dispatch Target	Regulation costs - raise/lower	$\frac{\textit{System requirement (metric)}}{\textit{Regulation eneblement amount}}$	FI is a non-linear function based on ACE and ACE Integral. This aligns most closely with the performance metric used in the current causer pays process. System requirement is the maximum (or minimum) of net system error over the DI.
2	Smoothed Hz	$\frac{\textit{Unit sumproduct}}{\textit{gross sumproduct}}$	Dispatch Target	Regulation costs - raise/lower	$\frac{\textit{System requirement (metric)}}{\textit{Regulation eneblement amount}}$	Smoothed frequency raw Hz filtered at a time constant of 35 sec. It approximates the need for regulation response. System requirement is the same as Scenario 1
3	Combined Hz (Raw Hz plus smoothed Hz)	$\frac{\textit{Unit sumproduct}}{\textit{gross sumproduct}}$	Dispatch Target	Regulation costs - raise/lower	$\frac{\textit{System requirement (metric)}}{\textit{Regulation eneblement amount}}$	This base scenario uses a combination of raw frequency (4 seconds on the mainland) and smoothed frequency, to value instantaneous and sustained frequency response. System requirement is the same as Scenario 1
4	Gross dispatch error	$\frac{\textit{Unit sumproduct}}{\textit{gross sumproduct}}$	Dispatch Target	Regulation costs - raise/lower	$\frac{\textit{System requirement (metric)}}{\textit{Regulation eneblement amount}}$	The *'Unit sumproduct'* is the unit MW deviation and the sign (expressed as -1 or +1) of the gross system error. *'System requirement'* is the maximum (or minimum) gross system error over the DI, which is significantly larger than the other 3 scenarios which use net system error

[8] Note that *sumproduct* is the product of metric x MW deviation summed over the DI.

Table 4-2: Procedure elements and identifiers for preferred base scenario variations

ID	Performance metric	Contribution factor	Reference trajectory	Financial weighting	Scaling	Commentary
3.1	Combined Hz plus frequency offset	$\frac{\textit{Unit sumproduct}}{\textit{gross\_sumproduct}}$	Dispatch Target	Regulation costs - raise/lower	$\frac{\textit{System requirement (metric)}}{\textit{Regulation eneblement amount}}$	This variation of scenario 3 includes a frequency offset to account for time error correction by AEMO. AEMO would need to signal to market participants when this offset would apply which may introduce complexities in relation to AGC and related communications systems. CF is normalised by dividing by the gross aggregate sumproduct
3.2	Combined Hz	$\frac{\textit{Unit sumproduct}}{\textit{gross\_sumproduct}}$	Target + regulation component	Regulation costs - raise/lower	$\frac{\textit{System requirement (metric)}}{\textit{Regulation eneblement amount}}$	This variation of scenario 3 includes the regulation component in the reference trajectory. CF is normalised by dividing by the gross aggregate sumproduct
3.3	Combined Hz	*Unit sumproduct only*	Dispatch Target	Regulation price – raise/lower	*Operators constant only*	This variation of scenario 3 is based on the IES/AEC/ARENA process for the implementation of frequency deviation pricing (ref). There is no normalisation of the CF
3.4	Relative performance	$\frac{\textit{Unit sumproduct}}{\textit{net\_sumproduct}}$	Dispatch Target	Regulation price – raise/lower	$\frac{\textit{agg net-sumproduct(di)}}{\textit{agg net sumproduct(lt)}}$	Scales amounts by short-term performance and long-term performance of the net error. Payments are scaled up if short-term performance exceeds long-term performance. An extra scaling parameter may be required to adjust the turnover (as in Scenario 3.3). CF is normalised by dividing by the aggregate net sumproduct

Table 4-3: Scaling factor scenarios

Scenario	Contribution factor (CF)	Scaling value - k	units	Ex ante / Ex post	Comment
3.8	$\frac{sumproduct(u, di)}{AGG(sumproduct(\dots))}$	$GrossError$	MW	Ex-post, as CF denominator and k are not known in advance	A variation on base Scenario 3. This scales the settlement amounts by the aggregate actual work done by plant deviations.
3.14	$sumproduct(u, di)$	$\frac{1}{75\ x\ freqref}$	1/Hz	Ex-ante, as k is effectively constant	This is a variation on Scenario 3.3, with the Operators Constant k defined by a measurable formula (for k). The reference frequency could be its standard deviation (currently 25 mHz) or the MPFR dead band (15 mHz). This is a reference frequency against which unit deviations are valued at the relevant deviation price. This scenario uses 15 mHz. Further analysis of this approach might suggest an alternative frequency reference.
3.15	$\frac{sumproduct(u, di)}{AGG(sumproduct(\dots))}$	$GrossError - measured\ AGC\ reg\ at\ this\ time$	MW	Ex-post, as CF denominator and k are not known in advance	A variation on Scenario 3.8 with AGC regulation included in the reference trajectory for AGC regulation units. This effectively removes AGC regulation from most of the performance incentive payments. The scaling factor is adjusted down by deducting the measured AGC regulation amount from the gross error.

5. Short Sample Period Results

5.1. Overview

After an initial filtering of options, the short sample period analysis was used to help understand and further narrow down options by modelling settlement outcomes and inputs to settlement outcomes. The short sample period analysis used SCADA and market data for the two weeks from 6 September 2021.

Several sets of analyses were performed:

- Analysis of performance factors using a narrowed down set of metrics and with other elements standardised. Regulation cost recovery was not included and performance was reported raw over the two weeks, without financial weighting or additional scaling. Scaling was standardised to allow comparison of the relative size of components. The four-performance metrics chosen distinguished our base scenarios.
- Analysis of complete settlement for both performance and regulation cost recovery, carried out in three steps, with each step being informed by the previous. Details are set out in e following sub-sections.

Reporting covered:

- A high-level comparison across scenarios
- Settlement outcomes over the two weeks for Raise and Lower, and for gross (DI level) and net (billing period level) outcomes
- A further analysis by fuel type for greater understanding of impact. However, we have only provided charting for the equivalent long term (11 month) analysis

5.2. High-level results for performance factors over two weeks

5.2.1. Description

The full set of summary results for the four base case scenarios is presented in Figure 6 following. The analysis spans a period of two weeks from 6 September 2021. The Raise charts are on the left of each scenario chart and Lower on the right.

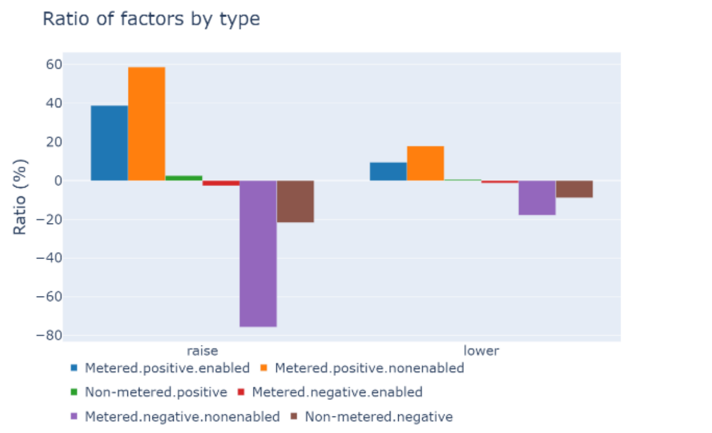
The characteristics distinguished in the charts are combinations of:

- metered and non-metered units;
- enabled and non-enabled units; and
- positive (receiving) or negative (paying) units.

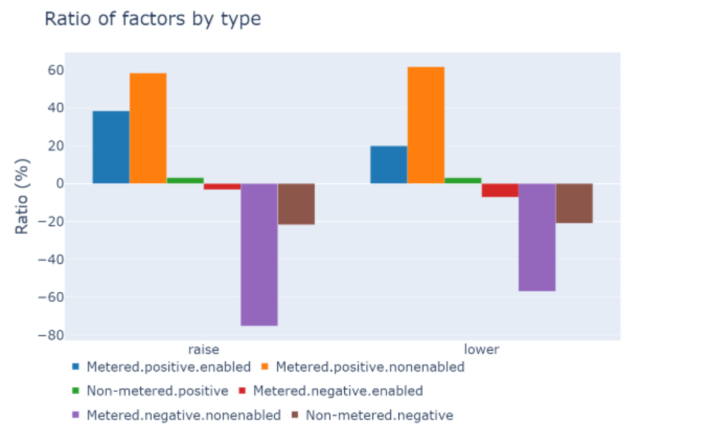
Non-metered units have no enablement option.

Figure 6: Summary performance factor results for Base Scenarios

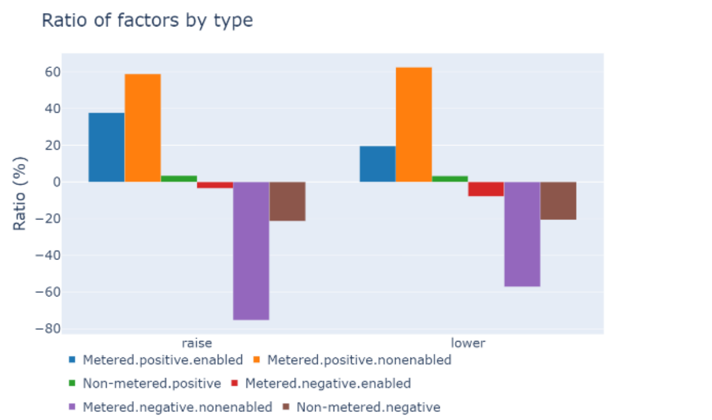
Scenario 1: Frequency Indicator (FI)



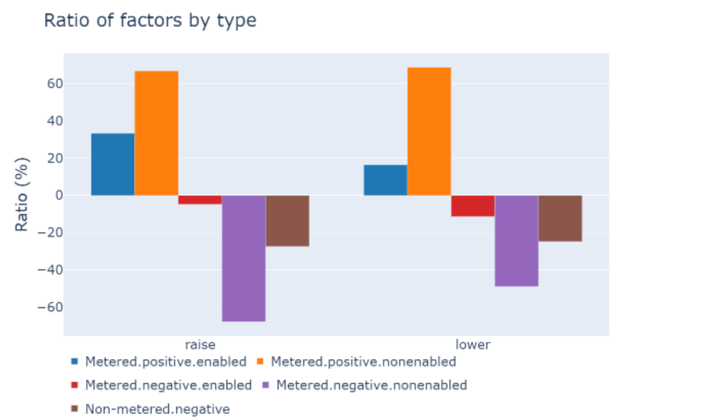
Scenario 2: Smoothed Hz



Scenario 3: Combined Hz



Scenario 4: Gross system error



The vertical axis is in percentages of the total raise positive factors, so the blue, orange and green bars on the left hand (Raise) chart sum to 100. All other bars are scaled in relation to these. So a comparison between the charts can only draw conclusions on the shares, not absolute values. Scaling will be considered under a separate set of analyses.

Using Scenario 3 as an example, we observe that enabled and non-enabled units make up most of the combined PFR and secondary response, as expected, although there is a small non-metered component; forecast error is less dominant than it used to be as a cause of frequency deviations. This is borne out by the metered negative factors (in purple) dominating the non-metered (brown) residual. Note that there are some net negative enabled performers as well, which ought to be surprising as enabled units are paid to perform.

The Lower chart on the right-hand side of Scenario 3 is broadly similar to that of Raise on the left. It seems somewhat surprising that the relative size of the performance measures is so similar. From these results, its seems that differences in settlement outcomes between Raise and Lower will likely be due mostly to differences in financial weights. However, there are some differences in other scenarios as we note in the following sub-section.

5.2.2. Comparison between scenarios

The most striking observation from the charts of Figure 6 is the similarity in relative shares between groups across all Scenarios. Apart from minor variation in the relative shares we observe the following:

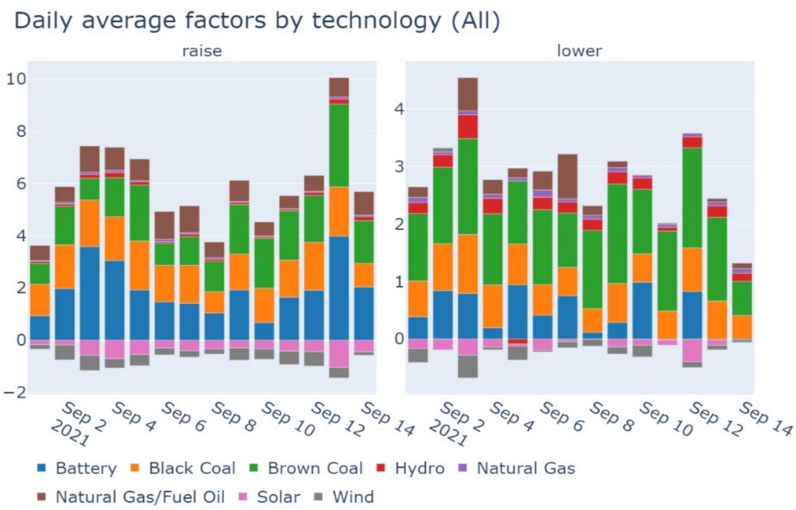
- There is no non-metered positive performance under the gross system error metric of Scenario 4. This arises from the definition of gross system error, which takes the residual deviations as being always bad.
- Under all scenarios, the positive metered performance and negative non-metered performance suggest that the management of Lower deviations is easier than for Raise, as would be expected. However, somewhat surprising is that non-metered performance is similar for lower and Raise in all but Scenario 1.
- In Scenario 1 which uses the AGC FI metric, the Lower performance measures are significantly smaller than for Raise and much small than for Lower in other scenarios. We have not investigated the cause of this in detail, but the FI metric differs from the others by having:
    - a significant dead band;
    - frequency offsets driven by accumulated time error.

  We note that the system historically has been responding to AGC regulation driven by the FI measure; not so for the other metrics. When the incentive is implemented and participants begin to respond, these out-turns may look different.

5.2.3. Daily Analysis by Fuel

An analysis broken down by day and by technology for all scheduled units for Scenario 3 is shown in the chart below. Because the residual is excluded in this chart, the factors do not balance. There is a wide spread of technology contributions presumably because of MPFR. Apart from a small negative factor from hydro on one day (possibly a single unit not conforming to schedule on ramp up), the consistently negative factors are from wind and hydro, as would be expected.

Figure 7: Daily average performance factors by technology for Scenario 3



The team has reviewed other detailed results, including results at the 5-minute level which have been set up for review on NEOPoint.

5.2.4. Conclusion

The broad similarities observable from this analysis does not imply that these metrics are equivalent. There are many other factors to consider when deciding on a satisfactory metric, such as the real time incentives that the metric offers for good performance.

5.3. Overview of settlement analysis

5.3.1. Introduction

The remaining results use the base scenario performance factors presented earlier and add financial weighting and scaling to deliver a full settlement analysis for the proposed settlement arrangement. As the AEMC Rule also make prevision for the recovery of AGC regulation costs, these payment and costs are included in the settlement results. As the Rule is likely to distinguish Raise and Lower, separate results are provided for these cases.

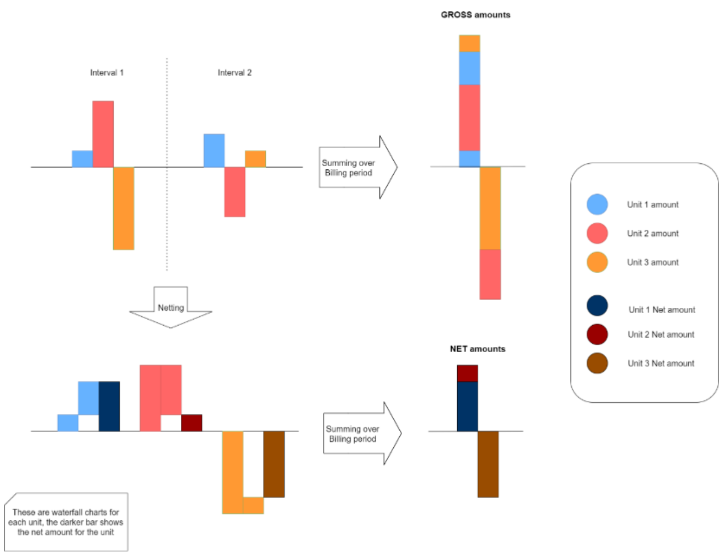
Results are mostly in the form of high-level summaries, but breakdowns by fuel/technology are also provided for the longer period case, together with additional explanatory charts and analysis. Reporting of the results is split as follows:

- Four base scenarios (1 to 4) over two weeks of data from 6 September 2021
- Four variations on Scenario 3 (3.1 to 3.4) for the same two weeks spanning specific options on some procedure elements)
- Three final variations based on Scenario 3 and scaling similar to the outcome of Scenario 4 over the same two weeks, spanning three different scaling options (Scenarios 3.8, 3.14 and 3.15). We performed an initial analysis over two weeks and repeated them as final variations spanning a period of 47 weeks from 1 February 2021 to 26 December 2021)
- Results broken down by fuel/technology for the three final variations spanning a period of 47 weeks from 1 Feb 2021 (ref).

5.3.2. Gross and net results

The results are also presented in two forms to aid understanding: gross and net. This distinction is illustrated in the diagram below.

Figure 8: Distinction between gross (top) and net (bottom) results



As shown in the top part of the diagram, the *gross* amounts are simply the *separate* summations of the positive and negatives. This corresponds to what is happening with settlement calculations at the metered 5-minute level.

In the bottom part, the *net* amounts are the *algebraic* summations of the positive and negatives, which can cancel each other in whole or in part. This corresponds to what is happening with settlement calculations at the billing level, when payments are made.

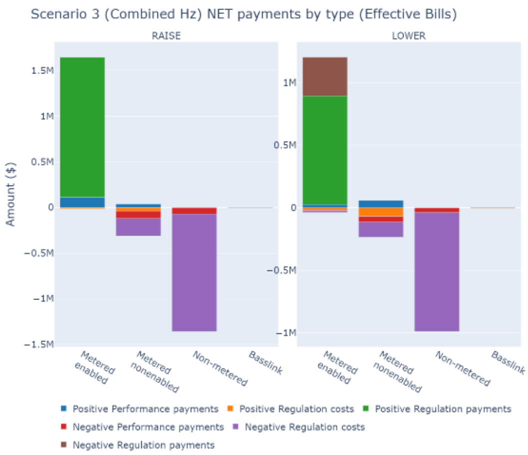
In this example, the top gross result is straightforward. In the bottom net analysis, unit 1 (light blue) has two positive amounts so the net (dark blue) is the simple sum and also positive. Unit 2 (red) has a positive first interval amount and negative and somewhat smaller second. The net summation is still positive but much smaller than the gross. Unit 3 (orange) is a large negative followed by a small negative, still giving a net negative.

This distinction helps understand why net amounts, which affect total bills, are so much smaller than amounts calculated at the DI level. The difference is due to the performance variability at the unit level and in total.

5.3.3. Example of a settlement summary

The analysis includes both AGC regulation and performance incentive payments and spans a period of two weeks from 1 September 2019. Following is a brief description of the content of the charts based on the example of Scenario 3 which is reproduced below. The chart shows Raise on the left and Lower on the right.

Figure 9: Base scenario full settlement results over two weeks



The broad categories distinguished in this net settlement chart are:

- metered enabled (for AGC regulation);
- metered non-enabled;
- non-metered; and
- Basslink.

The stacked bars contain the accumulated settlement amounts for payments (to) and costs (from) for both AGC regulation enablement and performance. Regulation payments are made to enabled units, and costs to negative performers. We do not in this analysis distinguish "used" and "not used" enablement volumes; the "not used" likely to be allocated according to a longer run measure determined from historical outcomes. However, the long-run settlement outcome should be similar. The following distinctions can be enlightening.

- Sometimes a unit can have a positive factor and payment for performance at the reported level, but still attract a charge for regulation because of periods of negative performance, even within a DI or over longer periods. These regulation costs are called "positive regulation costs" while the more dominant cases are called "negative regulation costs".
- The reverse can also occur: a unit's performance may be negative and so attract a charge, but it could still be paid for enablement. These are called "negative regulation payments" while the more dominant cases are called "positive regulation payments". A unit that persistently operates with negative regulation payments is not performing well when AGC enabled.

In this example the AGC regulation payments (green) and payments (purple) are the largest elements. Regulation payments of course go to enabled units while costs are shared between metered non-enabled units but mostly to the non-metered residual. Positive regulation costs and negative regulation payments are relatively small but present. The Lower chart is broadly like the Raise in total size, largely because regulation costs and payments dominate.

5.4. High level results for four base scenarios over two weeks

5.4.1. Description

The full set of settlement summary results for the four base case scenarios is presented in Figure 37 in APPENDIX E.. The elements of the four base scenarios are summarised in Table 4-1. The difference in these scenarios is the metric used and the scaling in Scenario 4:

- Scenario 1: Frequency Indicator metric (FI). The scaling factor is a net system error measure.
- Scenario 2: Smoothed Hertz metric. The scaling factor is a net system error measure.
- Scenario 3: Combined Hertz metric. The scaling factor is a net system error measure.
- Scenario 4: Gross System Error metric, implemented as the sum of unit error terms classified as positive or negative depending on the direction of gross system error. Expressed as a sumproduct, the product (at the measurement interval) is of the unit deviation and the sign (expressed as -1 or +1) of the gross system error. This delivers a performance measure where the only system -wide influence is on direction and not amount. The scaling factor is effectively a gross system error measure.

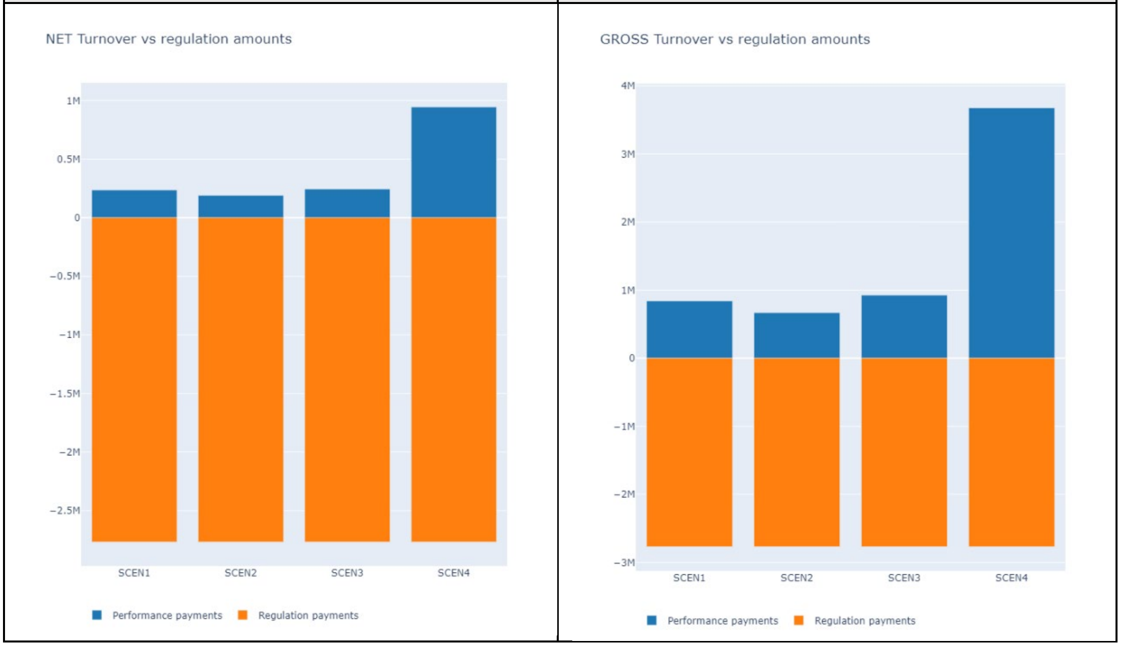
Because the Contribution Factor is normalised in each of these Scenarios and the financial weight is common, the only factor driving gross turnover difference is the scaling factor.

Further, the financial weight in these scenarios was regulation cost (Raise or Lower) and the denominator in the scaling is the reg gelation enablement amount (EA) in each scenario the effective financial weight is regulation price (Raise or Lower) and the numerator of the scaling factor in the table. This simplification was only appreciated after these runs were done so we do did not amend the table.

The results showing the four scenarios showing performance turnover (in red (in blue) and AGC regulation enablement turnover (in red) in red, presented as both net and gross, are displayed in the figure below.

Figure 10: Net and gross turnover for performance an AGC regulation for base scenarios

Net turnover in S	Gross turnover in $



5.4.2. Commentary

Features shared by all these results are:

- performance payments and costs are very much less than regulation payments and costs in Scenarios 1-3. However, the performance payments under Scenario 4 are much larger than for the other scenarios.
- the net results are very much lower than the gross in all scenarios i.e. netting is significant with many units having both positive (blue) and negative (red) performance at various times as shown in the gross, especially for metered non-enabled units.

There is an argument that scaling by gross system dispatch error instead of net system error or AGC regulation enablement would better reflect the MW requirement for both PFR and secondary response working together. This is despite knowing that gross dispatch error is in part simply dispatch errors in the metered unit population offsetting each other. The proposed incentives, together with better tuning of AGC control and the full participation of all units in mandatory PFR should minimise these uncontrolled errors with more control on the system.

However, the gross system error option as a system performance is problematic. While when normalised it gives broadly similar outcomes to a frequency-based metrics, it's incentive properties are poor. Consider two cases.

- If implemented directly as a performance metric, it oscillates in a binary fashion between one direction and the other, as is evident in Figure 1
- If implemented as a sign only indicator for unit deviations, a very similar and explicit binary incentive also applies.

A binary or effectively binary performance metric is problematic because it does not scale to instantaneous system need. The scaling factor does provide some moderation, but it is not proportional. For example, gross system error has an effective floor, so that total cash payments could be large in relation to desired performance when the instantaneous requirement is low. This could encourage vigorous but unhelpful responses at these times.

5.4.3. Conclusion

The much larger gross system error relative to net system explains why the Scenario 4 performance payments are so much larger than the payments driven by the net-focussed scaling of Scenarios 1-3. This suggests that the scaling approach will be a critical element of the incentive procedure. Gross system error or something with a similar outcome could be a possible scaling factor to achieve effective outcomes. Three scaling variations along these lines were initially analysed over two weeks and same scenarios were re-run and reported on over 11 months (47 weeks) in Section 6.

On the other hand, our analysis of performance metrics has concluded that the metric:

- should be frequency-based to reflect that the objective of the incentive is frequency control; and
- should provide proportionate instantaneous incentives through the metric.

5.5. High level results for four variations on Scenario 3 over two weeks

5.5.1. Description

The full set of settlement summary results for the four variations of Scenario 3 is presented in Figure 38 in APPENDIX E. The main elements of these scenarios are summarised in Table 4-2. Scenario 3 was chosen as the base scenario, not from the settlement results of Section 5.4 but from the consideration of viable metric options discussed in Appendix Section B.3. In summary Scenario 3 (which uses a combined Hz metric) was chosen over other metrics from the base scenarios because:

- being a frequency-based metric, it most directly targets the instantaneous system need for frequency control;
- it targets both primary and secondary response requirements;
- it is a metric that is proportionate to system needs; and
- it is transparent and easily calculated by participants using local measurements

Variations on Scenario 3 were chosen from other issues identified in the initial investigations as set out below.

- Scenario 3.1: Includes the time error correction frequency offset as determined by the AGC.
- Scenario 3.2: Includes each unit's AGC regulation trajectory in its reference trajectory
- Scenario 3.3: Implements an FDP approach as proposed by the AEC's ARENA-funded project (ref)
- Scenario 3.4: Implements "relative performance" scaling, where amounts are scaled by DI level performance factors relative to a long run value.

Further detail of these variations is provided in the Scenario-specific sections following.

The proper basis for comparison for each of these scenarios is with base Scenario 3 rather than between them.

5.5.2. Commentary on Scenario 3.1 – Inclusion of Hz offset

The clearest difference between Scenario 3.1 and Scenario 3 is that the non-enabled metered group show a poorer response in 3.1; greater variability in the gross and a more negative net response. The reason is that they tend to be responding to the MPFR requirement only which has no time error correction, so there is a greater mismatch between performance under a metric that accounts for time error correction and relative to performance under a metric closer to how they were operating at the time of the analysis current performance. This difference is to be expected when based on historical data.

There is an over-riding AEMC and AEMO policy not to use AGC signals in any incentive arrangement as AGC regulation settings may change. Nevertheless, it remains true that some form of time error correction would:

- allow non-enabled units to contribute to time error r, perhaps more effectively than the AGC reg system which struggles with time error correction because of its dead band setting; and

- avoiding the risk and inefficiencies of non-enabled units working against the time error correction of the AGC reg system.

Section B.3 canvasses an approach whereby a time correction compared is added to the performance metric. This can be made to operate in a smoother, more continuous fashion than the current AGC, resulting in smaller offsets and therefore tighter frequency control. We commend this in our recommendations for further consideration by AEMO at the design and consultation stage.

5.5.3. Commentary on Scenario 3.2 – Including reg signal in reference trajectory

Because AGC regulation is a major provider of both regulation and PFR, Scenario 3.2 is a significant change from Scenario 3. We observe the following:

- The system turnover is not changed because a fixed dollar amount determined by the financial weighting and scaling factors is still allocated, but in a different way.
- Metered enabled units display far more negative performance. Enabled units are more negative because most of the good performance was part of the reg signal, the benefit of which has been removed.

In summary, while some issues from this change are evident, there would need to be a relatively larger performance element to draw clear any conclusions. However, the lack of any change in the payment turnover even though large group of providers are paid less, appears problematic as it seems to deliver a windfall to non-enabled units and not relieve those who pay. As a minimum, this option should also attempt to net out the contribution of enabled units to from the scaling factor.

5.5.4. Commentary on Scenario 3.3 – FDP – IES-AEC variation

This option delivers a much larger performance element than the base Scenario 3 and therefore higher total payments. This is due to a different and more aggressive scaling arrangement than the base scenario including an operator's constant of 2. The performance payments to (blue) and from (red) are prominent. Specifically, we observe:

- Metered enabled units in total are the largest net performers, with little difference between net and gross. As expected, they are relatively good performers.
- Relatively large positive and negative performance in the metered non-enabled group in the gross collapse to much smaller levels in the net. This can be attributed to their relative lack of control over the period of the sample, reflecting a lack of performance incentive (beyond MPFR).
- The residual and a sub-set of metered units share the performance costs. Some metered non-enabled units are net positive performers.

In summary, because this variation happens to have the largest scaling of the variations in this set, we can see more clearly what is happening with performance payments. Needed is a comparison with similarly scaled variations of other options and a firmer basis on which to determine the scaling.

5.5.5. Commentary on Scenario 3.4

This variation also had different effective scaling, not form a different scaling factor but from the use of aggregate net performance factors for normalising the contribution factor. This delivers a larger scaling overall. This option was not considered further as the scaling could not be justified.

5.5.6. Conclusion

These variations on Scenario 3 confirmed the finding of the base scenario analyses of the importance of a robust scaling approach supporting both PFR and regulation. This scaling will be larger than the implied scaling in the Draft Rule, which proposed Regulation Requirement (assumed to be the regulation requirement as measured by the Frequency Indicator, FI.

Scenario 3.1 included an AGC-generated time error correction offset. The run highlighted that this metric modification made a difference. However, use this AGC variable is problematic because of the sound policy requirement to avoid direct interaction with the AGC. Therefore, we have proposed an alternative approach in Section B.5 that is smoother but does not reference the AGC. For this reason, we did no further runs on this issue.

Inclusion of the AGC reg trajectory in the reference trajectory in Scenario 3.2 is of interest to participants but, under the contribution factor approach, the performance turnover remains unchanged. Regulation enablement charges still dominates, however. This option was examined further with improved scaling as Scenario 3.15 in both short period and long period runs, as reported in Section 6.

Scenario 3.3 is of potential interest but it's scaling is arbitrary and needs to be revised with a more robust scaling logic. This is done in later short and long period analyses as Scenario 3.14.

Scenario 3.4 uses an aggregated net measure as a normalisation factor which inflates turnover relative to other scenarios. Because we have concluded that a gross approach is more appropriate, this form of normalisation cannot be defended and this concept was not pursued further.

All these scenarios also confirmed the importance of a robust, defendable and implementable scaling logic in any arrangement.

6. Long Sample Period Results

6.1. Description

Based on initial investigations and the short period scenario runs, we developed a set of scaling options that were both measurable and reflective of the defendable gross error scaling approach that our analysis supported.

Table 6-1: Element options for used for long period Scenario runs

Element	Option
Performance Metric	Combined Hz
Reference Trajectory	Target to target; no AGC reg except for Scen. 3.15
Classification	Separate Raise and Lower; performance classification made at measurement level
Unit Aggregation	Unit level settlement
Residual Calculation	Top down / Energy balance
Financial Weight	Regulation price
Scaling	Several options tried with two preferred. More analysis required. See below.
Regulation Enablement Cost Distribution	Used to negative factors, unused to negative factors average over an historic period.

The scaling and related elements of the settlement formula are summarised in Table 4-3. Scenario 4 is the closest reference scenario reference scenario for these cases although these scenarios can reasonably be compared to each other. They key differences are:

- Scenario 3.8: Gross system error scaling. The MW requirement for Raise is the maximum (or minimum if Lower) measure MW gross error as measured over the DI.

- Scenario 3.14: Hz spread scaling. Under the FDP approach, the scaling is an inverse frequency value. This was set at an inverse MPFR dead band value, meaning that the effective price at this deviation is the same as the financial weight. This scaling can be related to a gross error type of scaling with some algebraic manipulation: see APPENDIX C

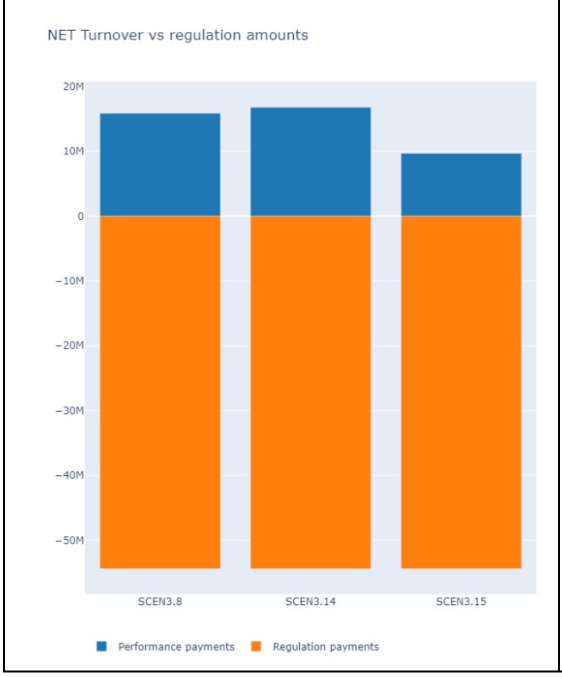
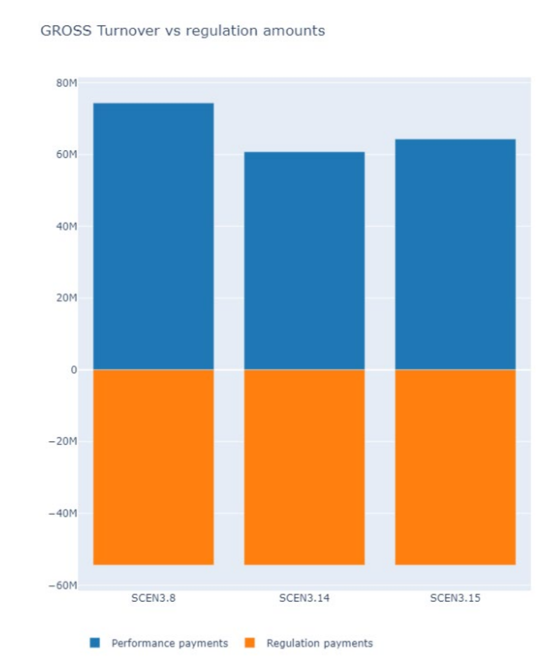
- Scenario 3.15. Scenario 3.8 but with AGC reg trajectory included in the reference trajectory. The total AGC reg trajectory being netted off the gross error to determine the scaling amount. This is done because most of the payments to enabled units would be removed, even though enabled units are still doing significant work.

6.2. Comparative overview of scenario settlement results

Net and gross results comparing the three scaling scenarios in this set are shown in Figure 11 following. In these charts, Raise and Lower are combined. The charts include both AGC regulation and performance elements. Presented are the net and gross turnover figures.

Note that the vertical scales are different in the net and gross charts; net total turnover is significantly larger in the gross case relative to the net due to differences in performance turnover. The regulation settlement turnover is the same between scenarios and between net and gross; it is provided for comparison.

Figure 11: Performance and regulation settlement comparison between scaling scenarios

Net turnover in S	Gross turnover in $
	

From this chart and in comparison, to the 3.X results we observe the following:

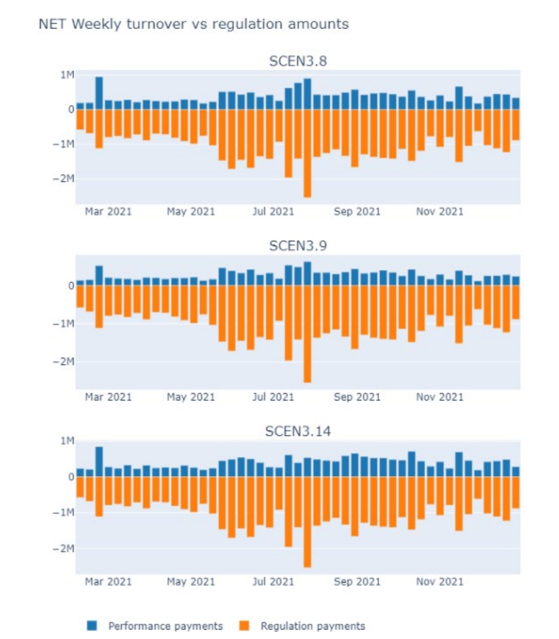
- The performance turnover is much larger in all these cases than any of the base scenarios except Scenario 4, as desired.
- The performance turnovers in and 3.8 and 3.14 are broadly similar. Scenario 3.15 is less than the other because of the removal of AGC reg and the work that it does. However, it is notable that the gross turnover is similar to 3.8 and 3.14; the higher level of shrinkage in the net is likely due to the poorer performance under the metric of the non-enabled units, relative to the mix of enabled and non-enabled units participating in the other two.

- There appears to be relative less netting out in Scenario 3.14 relative to the others; the gross performance turnover is the lowest but the net the highest. This suggests that this settlement approach measures performance a little more sharply than the others. This warrants further exploration.

6.3. Weekly settlement amounts over 47 weeks

Figure 12 below shows net and gross weekly turnover and regulation amounts. Raise and Lower are combined in these charts. The results may be affected by the ongoing rollout of units under the MPFR rule, although the major providers would likely have been operating by the start date of this analysis. With these reservations, some observations follow.

Figure 12: Performance v regulation enablement costs by week

Net Weekly Turnover and Regulation Amounts in $	Net Weekly Turnover and Regulation Amounts in $
	

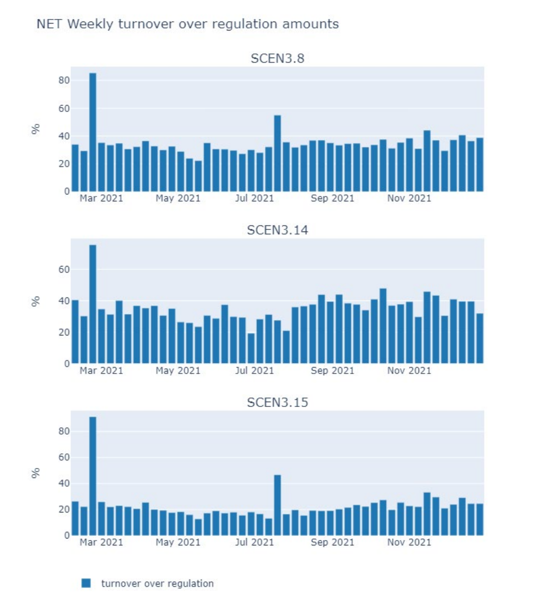
- There is a clear relationship between gross payments (blue) and regulation costs (orange), as expected from the financial weighting choice of regulation prices.

- There also appears to be a seasonal relationship, with higher costs and associated turnover tending to move higher through the winter period. This may be due in part to more variable condition driving the renewables sector.

- There are a few weekly outliers. We can speculate that these are due to some incident in the system. However, on a weekly settlement basis the variations are not particularly notable.
- The difference between net and gross performance payments is strongly evident, as would be expected.

We can discern some further interesting differences in the charts in Figure 13 below. The charts show the ratio of performance payments to regulation payment by week, for net and gross cases. The regulation price is in both numerator and denominator, so its impact is largely removed. The variation that remains can largely be ascribed to variations in the scaling logic including whether or not a “sharing“ type of contribution factor approach has been used (3.8) or not (3.14). We make the following observations:

Figure 13: Ratio of performance and regulation turnover by week

Ratio of net performance to turnover payments by week	Ratio of gross performance to turnover payments by week
	

- As expected, the average ratios reflect the observation made in the summary results on the relative rankings in the gross and how they change when adjusted to the net.
- There appears to be significantly more weekly variation in 3.14 than in the other two. This is true for both the net and gross cases
- There is an outlier in March in all cases, likely traceable to a system incident

The difference in weekly variability is interesting. The less volatile options would track regulation prices closely and show less variability due to performance. The more volatile option (3.14) suggests greater variability due to performance. The impact of regulation price on all options should be similar. A weekly ratio would average out a lot of performance variation. It is likely that the variability would be great at the daily level and even stronger at the DI level. However, we have not yet examined this issue in any detail.

This observation suggests that there may be different types of incentives at the micro level under these two approaches, even though the settlement amounts overall are similar. We suggest a more detailed investigation during AEMO's implementation process before settling on a final scaling logic. This is included in our recommendations.

6.4. Summaries of full period settlement results

These charts are presented in Figure 39 of APPENDIX F.

While there are differences between these results in terms of total performance payments, the surprise is how similar they are given the quite different bases for each scenario.

For example, Scenario 3.8 scales by gross error, as measured by the maximum (or minimum) gross error above or below) zero deviation in a DI as measured at the time of settlement. Gross system error is typically many more than times the usual AGC enablement amount of about 200 MW. However, an investigation into revised AGC tuning might well consider whether regulation enablement and 4-sec dispatch should be targeted at something of the order of the gross dispatch error.

Scenario 3.14 has a quite different logic again, in that is not based directly on sharing out a dollar amount determined by a gross MW volume and the weighting price. Instead, it is driven by an FDP linked to the financial weighting. This approach appears to be somewhat less susceptible to shrinkage when moving to the net level. The explanation may be that it picks up less of the random noise in the system than the other options.

Scenario 3.15 shoes a much-diminished performance payments to enabled units, as intended and expected. Because of the higher levels of negative performance now assessed for enabled units, they also attract more enablement charges, thereby providing an indirect financial performance incentive to enabled units.

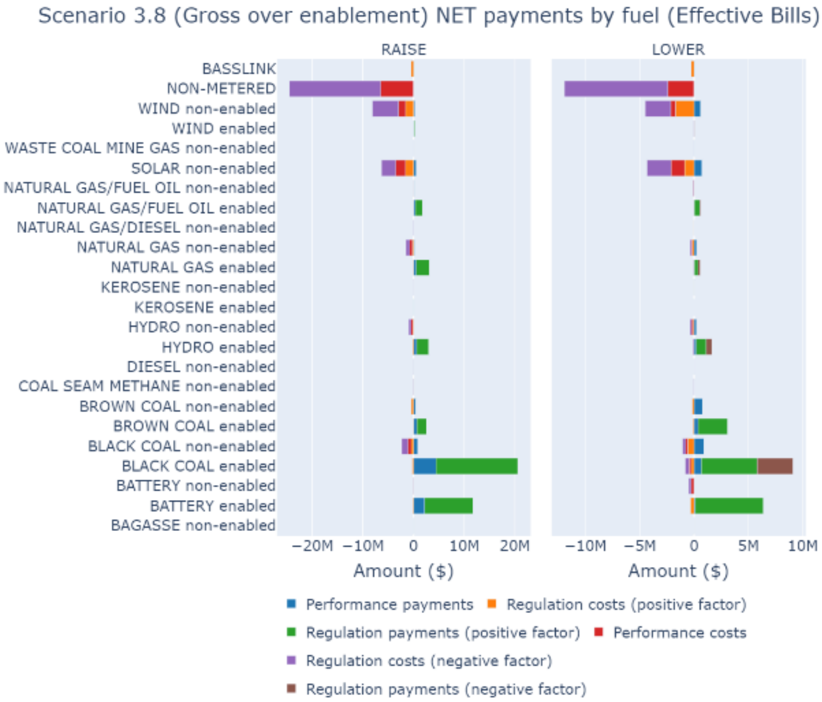
We conclude from that these all appear to deliver reasonable performance payment volumes based on justifiable criteria. The choice between them would be based on a lower-level analysis of the performance incentive offered in each case, which may be structured differently in different circumstances.

6.5. Full period settlement results by fuel/technology and enablement

6.5.1. Description

Gross and net charts for all three long period scenarios ae presented Figure 40 in APPENDIX F. An example chart for that Appendix is presented in Figure 14 following. It shows the net results for Scenario 3.8.

Figure 14: Example of Settlement by Fuel/Technology and Enablement



The chart presents gross payments (to) and costs (from/charged) for a range of fuel type/technology classes, including batteries. All these scheduled units are split into enabled for AGC regulation and not enabled. Unscheduled items are non-metered (the residual) and Basslink, which is the only external DC link and which, while it has some frequency control capability, would operate as an unscheduled generator and load in thebe ignored in the calculation of contribution factors in the proposed system. It is only included in the modelling to ensure totals are calculated accurately and all deviations are accounted for.

The chart is a “picture in time” and reflects a state where nearly all large thermal units have implemented MPFR, but not all, with some inverter-based units needing upgrading their software. It should be noted that thermal units operate with noticeably different regimes, for example the super-critical coal units Kogan and Millmerran are far less likely to be providing primary response due to heavy loading of the units and focus on steam flow over dispatch control, as opposed to older sub-critical and export exposed units like those in NSW.

The charts also show the allocation of AGC regulation payments (to) and the cost allocation (payment from). The following distinction is also made, which can be enlightening.

- Sometime a unit can have a positive factor and payment for performance at the reported level, but still attract a charge for regulation because of periods of negative performance, even within a DI.
- The reverse can also occur: a unit's performance may be negative and so attract a charge, but it could still be paid for enablement. A unit that persistently operates in this manner is not performing well when AGC enabled.

6.5.2. Commentary

The 80-20 rule is clearly operating with the chart above and the complete set for all three long-term scenarios of Figure 40, although largely reflecting the relative preponderance of fuel types in the system. We observe that:

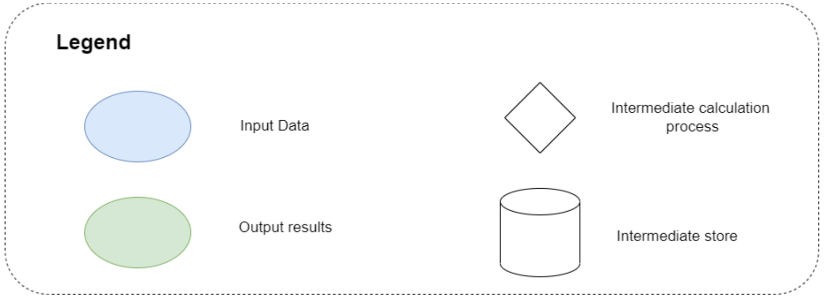
- While performance and regulation turnover are nearly equivalent at the gross level, performance payments at the net level are much less than enablement payments. In the example chart, green and purple are the largest elements.
- The top receivers of AGC regulation payments are black and brown coal and batteries (AGC enabled) as well as smaller amounts from gas, brown coal and hydro, and the payers are wind, solar and the residual.
- As would be expected, enabled units currently dominate performance payments, although non-enabled black coal is also represented.
- Performance payers would be renewables and the residual, following the enablement cost allocation as expected. However, performance incentives should drive up enablement payments which would flow through to the rest of the system.

6.5.3. Conclusions on fuel/technology settlement analysis

Under scenarios 3.8 and 3.14 there is a high correlation between providers of enablement and the likely recipients of payment under a performance incentive arrangement, as would be expected based on historical figures. The exception is non-enabled black coal, no doubt because of MPFR requirements. The positive performance payments are spread across all fuel types but they are not in total as large as black coal although wind and solar tend to be payers.

APPENDIX A. Calculation Elements

Figure 15 Diagrammatic representation of the calculation procedure



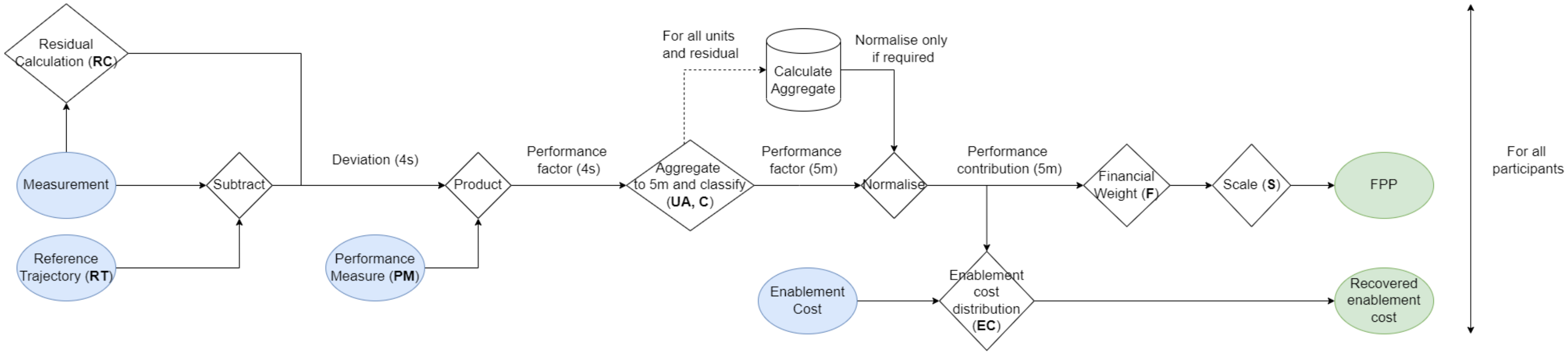


Figure 15 shows a diagrammatic representation of the complete calculation process with elements identified in XXX marked with bold text. A more detailed discussion of each element is provided below.

A.1 Performance Metric

A deviation from the reference trajectory can either be correct system error (good) or add to system error (bad). The performance metric is used to identify good and bad deviations and provide a weight representing the "importance" of a particular deviation. The sign-only group of performance metric will only provide a direction and not a magnitude of importance.

The period-by-period product of the performance metric and the unit's deviation results in the unit's 4sec performance factor. The 5-min performance factor (or just performance factor) is the aggregate (sum) of 4s performance factors within each 5-min period and according to the unit-aggregation process (See Section A.4). This performance factor can be normalised by the aggregate performance factor (See Section A.5 for a discussion on normalisation).

A.2 Reference Trajectory

The reference trajectory is the reference by which unit deviations are measured. The reference trajectory should represent the trajectory expected of the unit's output by the system operator. In the current causer pays system, the simple linear trajectory from target to target is used as the reference trajectory, the recent discussion is focused around including a droop or regulation component in the reference trajectory.

A.3 Classification

The choice of classification method and when to apply the classification is expected to make a significant difference to the results. Classification before aggregation (Options C1.X.X and CX.1.X) implies that each 4second performance factor is classified and then aggregated (See Section A.4) into 5min values, 1 per classification. Classification after aggregation (Options C2.X.X, CX.2.X) implies that 5min aggregate performance factor is calculated and then it is classified according to the chosen option. e.g., If the option selected for this element was C1.1.0, then there are potentially 4 5-min performance factors that gets calculated for each dispatch interval and unit (or portfolio): raise-positive, lower-positive, raise-negative, and lower-negative. Each of these classifications can have a separate financial weight applied to it. We will investigate the impact of these classifications and whether such complex classifications would result in more efficient outcomes.

The option that exists in the current causer pays system is C2.1.1 would result in 4 types of 5min factors. The benefit of using a simple option such as C2.0.0 is that there is only a single type of 5-min factor. The effective price applied to each 4-second deviation is the same regardless of any classifications. We note that AEMC staff plan to provide some policy guidance on the preferred approach to classification e.g., for reg enabled units.

A.4 Unit Aggregation

The unit aggregation element determines how 4sec performance factors (or performance measures) are aggregated to 5min performance factors. The unit aggregation (equivalent to no aggregation) is the simplest where 5min performance factors are calculated for each unit (i.e., each unit is considered separately). A portfolio aggregation is what is done in the current causer pays system where performance factors are calculated for portfolios of units; these portfolios are grouped into units owned and operated by the same commercial entity. Ideally, such a grouping should make no difference to the financial outcomes.

A.5 Financial Weight

The purpose of the financial weight is to represent real-world costs as they vary in different DIs. The financial weight should reflect the potential costs of provision of frequency response e.g., regulation prices/costs or energy price. Since a there is no market for PFR enablement, some proxy is required in that case. A cap or a floor (or other suitable modifier) in the financial weight may be required as prices and costs can become very high during tight conditions.

A key consideration in the financial weighting is whether to weight the contributions in aggregate, based on say the aggregate cost of regulation services or to weight the contributions individually based on some other metric e.g., a local energy price including MLFs.

A pricing system and a cost allocation system are very similar. By the choice of options, a pricing system can be made to return similar or similar but not identical results as a similarly optioned cost allocation system. A significant difference in this classification is how the objective of the system is perceived by stakeholders and operators. Under a cost-allocation system, a global cost is determined and allocated to all participants. Under a pricing system a price is determined for all deviations of a unit.

A.6 Scaling

The purpose of the scaling step is to increase or decrease the performance payments by what is perceived to be the requirement of the system, i.e., if there is a greater requirement for decentralised frequency response, then the performance payments should be scaled up and if there is a lesser requirement, the payments should be scaled down. The scaling parameter can also be identified ex-post (after the end of the dispatch interval) but better to keep it known in advance and adjusted ahead of time if recent trends suggest a change is desirable. It should be noted that, even if the scaling is a fixed value, such as in scenario E.14, it is necessary to vary the level of cashflows within the interval, and scenario E.14 does this by directly paying on the unit performance, which is the deviation * performance metric * scaling factor, and not calculating a normalised contribution factor, as a number between -1 and 1.

Another reason to include a scaling parameter is to gradually ramp up the payments/costs such that participants and AEMO can get used to the new payment mechanism and measure responses.

Other options that have been identified are using the fraction of responsive plant online in the system (option **S2**) or a constant that is calculated by the operator of the system (option **S3**).

A.7 Regulation Enablement Cost Distribution

The enablement market for regulation will continue to function and hence the method to distribute costs of enabled regulation needs to be specified. In the current procedure, all enablement costs are recovered from negative performers based on their negative contribution (option **EC2**). The draft rule specified that used regulation cost should be recovered from negative performers, but unused regulation cost should be recovered from all participants based on their share of generated or consumed energy (option **EC1**).

Another option is to allocate all such costs to the Residual on the basis that scheduled units re incentivised by a double-sided arrangement. By avoiding a single sided facility in this way, opportunities for portfolios not available to single units would be removed.

APPENDIX B. Initial investigations

B.1 Overview

In the following section we summarise and expand where necessary on the work done to narrow down the options for further investigation. There is one section for each element. Each section begins with a tabular summary, followed by a more detailed discussion of the analysis done performed and reasons for acceptance or rejection of an option. The section concludes with summary of the option or options to be carried forward.

This work led to sections describing the scenarios chosen for the short (two week) analysis and describing the results of that analysis. A discussion of some basis concepts follows.

B.2 Definitions of terms used in a dispatch error approach

The proposed new arrangement could use dispatch error measurements in several ways:

- as a performance metric; the measure of system need against which unit deviations are assessed; or
- as a method to scale the output when the metric might be based on some other measurable variable such as frequency.

The dispatch error approach looks at aggregate (system-level) deviations from reference trajectories. The following terms are defined (see sample chart in Figure 16).

System above response: the sum of all metered responses that are above the reference trajectories.

System below response error: the sum of all metered responses that are below the reference trajectories.

System net response or aggregate/net system/dispatch error: is the algebraic sum of the above and below responses, representing the effect of unmetered plant on the system aggregate response. (Figure 16)

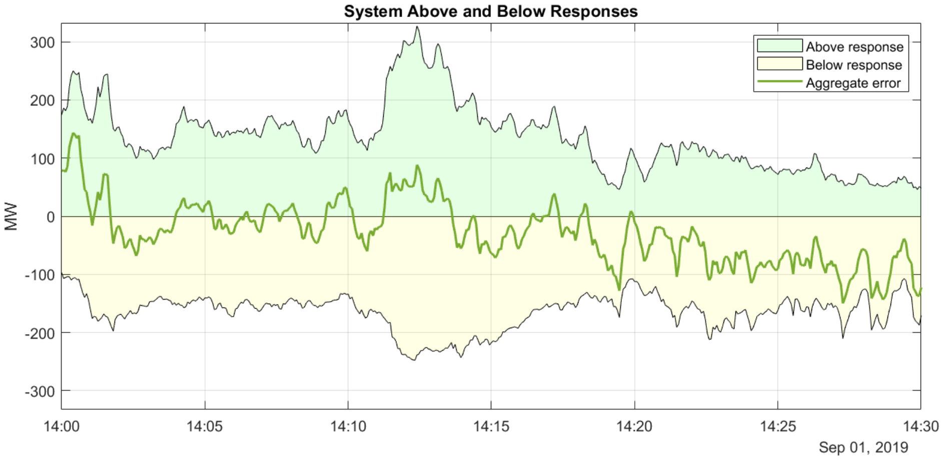
$$System\ net\ response = System\ above\ response + System\ below\ response$$

Majority response: the larger (in absolute magnitude) of either the system above and or below responses. Note that, by definition, the system net response is always in the same direction as the majority response. The majority response is referred to as the **gross dispatch error** or **gross system error,** shown by the red line in Figure 16.

Minority response the smaller (in absolute magnitude) of either of the system above or system below response errors.

Unmetered error correction: the portion of the majority response that corrects offsets for the unmetered elements (e.g., due to forecast error and load relief). Equivalent to the net system response.

Figure 16: Above, below, gross and net response



Metered error correction: the portion of the majority response that corrects for the metered elements (e.g. due to dispatch deviations). Equivalent to the majority response less the net system response.

Note the **net dispatch error** and **gross dispatch error** were considered as candidates for the performance metric and/or scaling factor as it would include both primary and secondary response. The hypothesis was that with tight dead band MPFR, a high aggregate droop capability on the system may absorb a significant dispatch error without a large frequency deviation. For example, for the same dispatch error, the frequency deviation may differ depending on the aggregate droop (PFR) capability. Without knowing at any one time the aggregate droop capability, by measuring the gross dispatch error rather than simply the frequency deviation, the primary response was hoped to be accounted for.

B.3 Choice of Performance metric

B.3.1 Overview

Table 6-2: Initial options and modifiers - performance metric

Base option or modifier	Identifier/ Reference	Discussion

Raw (4 second) Hz	PM1.X	• Captures fast acting requirements
		• Rejected (as sole option) for reasons below
		  ○ Very volatile
		  ○ Does not capture value of sustaining response required to control Hz
		• This option is used as one of the components of PM4.X

Smoothed Hz	PM2.X	• Captures need for sustained response • Rejected (as sole option) for reasons below

Base option or modifier	Identifier/ Reference	Discussion
		○ Does not fully capture value of short-term response • This option is used as one of the components of PM4.X
Hybrid – FI	PM3.X	• Drawbacks of this option ○ Cannot be locally measured ○ Creates hard link to AGC ○ Does not capture value of short-term response • This was used for Scenario 1 in the simulations and intended as a benchmark
Combined Hz and Smoothed Hz	PM4.X	• Not rejected • Used for Scenario 3 in the simulations
Area Control Area (ACE)	PM5.X	• Rejected as redundant for reasons below ○ Relies on AGC which is to be avoided as settings might change. ○ ACE is approximated as a multiple of frequency and so is very similar to Hz + time error offset (PM8.)
Net dispatch error	PM6.X	The direction of net error could be used to inform the positive direction of unit responses
Gross dispatch error	PM7.X	See above for discussion on direction of gross error • The gross error is a good indicator of the total response (not the net response) of units. • This was used for Scenario 4 in the simulations. • However, from the observed data, we see that a lot of metered response is balancing other metered responses, this indicates that the above assumption may not hold.
Metric adjusted by time error offset	PM9.X	• This option is rejected as a sole option (see discussion in PM2.X) • It could be one of the metrics under option PM4.X
Product function performance measure	PMX.0	• The product of desired response with actual response is a preferred because it ○ reflects both system need and unit response in a single measure ○ has a basis in control theory and practice
Sign-product function	PMX.1	• This option is rejected as sign-only metric would ○ value only the magnitude of the actual response and take no account of the magnitude of the desired response
Sign-dead band function	PMX.2	• This option is rejected because ○ Dead bands can distort behaviour by creating boundary effects

Base option or modifier	Identifier/ Reference	Discussion
		○ In any case units can implement their own dead bands
Correlation function	PMX.3	• This option is rejected because ○ The magnitude of response would not be valued ○ Correlations at 5-minute are very volatile and do not yield sensible results (p-values are high) • However, note that performance function data, when accumulated over a long period such as a billing week, provides useful covariance and correlation information.

Under the Draft Rule, the AEMC has determined that mandatory PFR (MPFR) is to endure. This opens several possible approaches to formulating a metric for PFR:

- Since units will have an MPFR delivery obligation under the MPFR Rule, one approach is simply to ask whether a unit is compliant with its obligations under the Rule or not. Ideally this would follow a philosophy that such matters be settled at the DI level.
- If a unit is assessed as compliant over a DI, it would receive a payment, presumably proportional to some sensible measure such as nominal obligation or measured performance.
- If assessed as non-compliant over a DI it receives nothing, except possibly a warning if non-compliance persists.
- Alternatively, a payment could be considered and structured as a performance incentive, to encourage performance to the mandated level and possibly to improve on that performance.
- In both cases, the payments should be consistent with the payments made for AGC regulation as these services, while distinct, overlap for much of the time.

The options we considered fall into one or sometimes both categories.

B.3.2 Yes/no evaluation within the DI

This approach seems simple and attractive but to implement it we need:

- a measurable threshold calculable within each to support a yes/no assessment; and
- a defendable size or performance-based metric on which to base a payment.

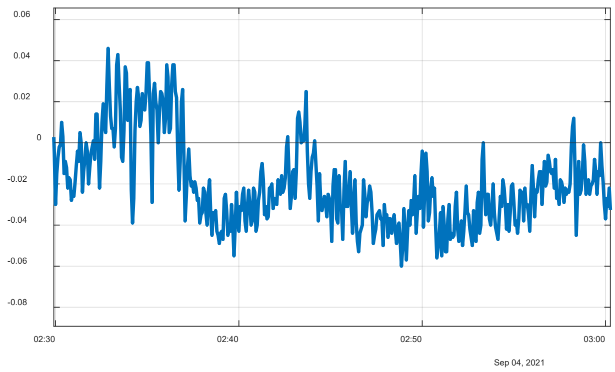
Thus we can't avoid the need for some form of continuous performance metric to make a Yes/No determination.

B.3.3 Correlation over 5 minutes v. correlation over the long term

APPENDIX D presents an analysis where we attempt to assess performance relative to frequency by calculating a correlation coefficient within a 5-minute DI. Our conclusion is that the sample in 5 minutes is too small to get a meaningful result. This was improved but still unsatisfactory at half an hour. Only over a long period of, say, of a billing week, could meaningful results be obtained.

Apart from the small sample size, another reason for this difficulty is the current behaviour of frequency as shown in the figure below. Over a long period, frequency tends to a zero mean because time error is controlled. However, over this half-hour period and many other periods of one or more DIs, frequency is substantially biased in one direction: in this case mostly down.

Figure 17: Half-hour sample of Frequency behaviour



B.3.4 Metric based on frequency

As frequency is the variable to be controlled, frequency deviation from its base level (50 Hz) is a natural choice for a metric because:

- It directly targets what the incentive system is attempting to do; namely, to control frequency
- It is easily implementable, measurable and understood from both centralised and local measurements
- It can be designed to target both primary and secondary frequency responses.

However, there are potential challenges. While the standard for SCADA communication in the NEM is a lag of no more than 6 seconds, for some sites, that lag is a lot longer. Indeed, 6 seconds is on the boundary of what would be acceptable. However, in the case of narrow band PFR, the main requirement at present is a sustained rather than instantaneous response.

Where a site has a large communication lag with AEMO, it could seek to upgrade its communications systems. One of the useful properties of a simple frequency metric is that its value is maximised when the lag is minimised.

We would also commend the concept of sites being able to send AEMO packaged frequency and unit MW high resolution data in a format to be agreed, and for an agreed and reasonable fee if necessary. The arrangement should be standardised as much as possible and be auditable.

B.3.5 PFR metric only

One option for secondary control incentives is to rely entirely on the raw frequency metric. According to this set of arguments, a lagged measure would:

- repeat the problems of FI the measure, such as running counter to frequency so that whole DIs are dropped; and
- pander to old technologies which are rapidly being superseded by batteries.

On the first point, we have outlined in Section xxx how primary and secondary controls can work together to produce a stable and sustained outcome. This can be done in a way that does not unnecessarily tie up the valuable fast response options needed to help protect against contingencies.

While it is likely that the fast response of batteries could well dominate the provision of these services as time goes on, this fact does not eliminate the lags in the rest of the system. If lags in the system remain, reliance on proportional control logic alone to eliminate error can lead to instability, as is well known in the theory and practice of control. Of course, AGC regulation would remain to provide a secondary service, but this would not meet one aim of the revised incentive arrangements, which is to improve the performance of regulation.

The objections to using a lagged/smoothed metric appear to be related to the nature of the Frequency Indictor (FI), the metric used by the current causer pays process. There seems to be a common misunderstanding that a metric that runs counter to frequency regularly' or even sometimes' is inherently bad, whereas a metric driving such a response is necessary to sustain an output without tying up valuable fast-response resources. The problem with FI running counter to frequency and penalising good fast response frequency performance was the absence of a PFR component in the metric to reward or charge fast moving units appropriately.

B.3.6 Frequency Indicator (FI) as a metric

Use of the current causer pays metric, FI, as a performance metric under the new arrangement presents the following challenges.

- AEMO's AGC system is subject to minor and major future changes, presenting risks to the new procedure as well as limiting AEMO's options for change.
- **T**he components of the FI metric are generated in real time within the AGC and can therefore only be known after the event. While it could conceivably be transferred to users at close to real time, it would be far less satisfactory than relying on local measurements if real time transparency (ex-ante operation) is an aim.

- Even after the event, the FI metric is difficult to reproduce in a way that makes it useful for local estimation.

- The metric is highly non-linear. Specifically, it has a relatively wide dead band, which is somewhat disguised by the signal being passed through a low-pass filter after the variable gain has been applied.

The virtue of FI is that it is a frequency-based metric (with time error correction) and directly derived from the AGC's assessment of the secondary control requirement. This is not sufficient to overcome its disadvantages, given that there are alternatives. Therefore, we propose to assess the FI metric, but as a reference only.

B.3.7 Smoothed frequency as a metric

We can be inspired by the theory and practice of control and consider a type of smoothed frequency response that is somewhat similar FI but bypasses its difficulties stated above. Below is a plot of smoothed frequency metric calculated according to a low-pass filter formula:

$$Smoothed_{Hz_{dev(t+1)}} = (1-a) \times Smoothed\ Hz_{dev(t)} - a \times Measured_{Hz(t)} \qquad (3)$$

where *delt* is the measurement interval and

$$a = \frac{delt}{Time\_Constant}$$

The formula for a digital version of continuous filter is slightly different, namely:

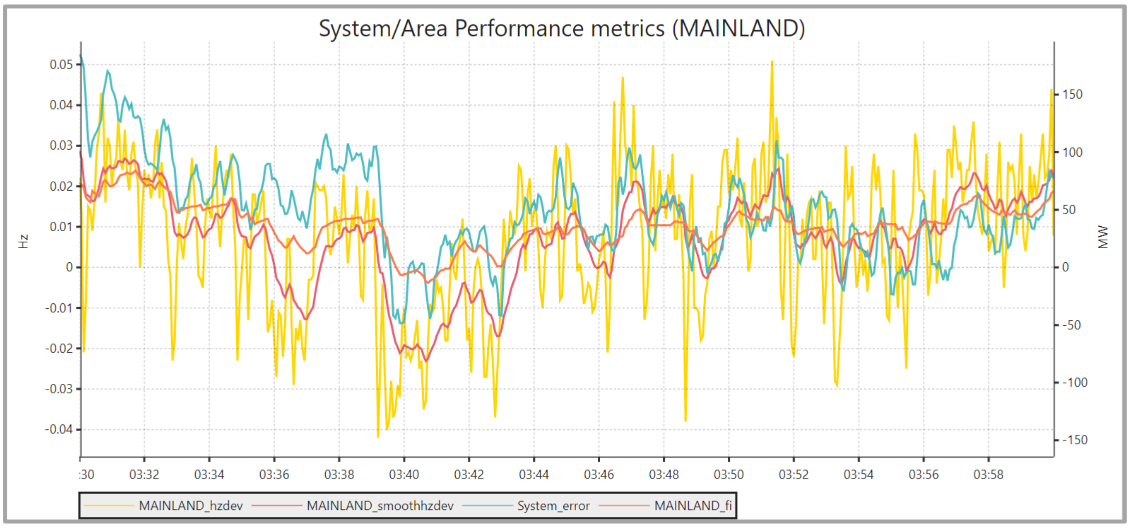
$$a = \frac{delt}{Time\_Constant + delt}$$

The AGC time constant using the first formula for a above is measured at about 34 or 35 seconds with a 4-second measurement interval. Using the second formula a time constant of 30 seconds gives an equivalent result. This may well be the default setting in the AGC.

The negative sign on the second term applies because the sign of the metric is the reverse of the sign on the frequency measurement.

Figure 18 following illustrates quite clearly the smoothing effect of the filter. (The raw frequency deviation is in yellow, and the smoothed version is in red in Hz on the left-hand scale). This compares with the system FI from the AGC in orange and the net system error in green (in MW on the right-hand scale). However, the intended effect is not only smoothing, but to sustain a response even when the frequency has returned to close to zero.

Figure 18: One hour sample of raw and smoothed frequency metrics



The advantages of this smoothed Hz metric are:

- It is easily calculated using local frequency measurements
- It meets a sustained response requirement when frequency returns to a small value
- In the form described above, it has no dead bands and so supports parties who may wish to operate at a tighter dead band than that required under MPFR.
- It can easily be combined with a raw frequency PFR metric if desired.

For these reasons it is a good candidate as a secondary metric.

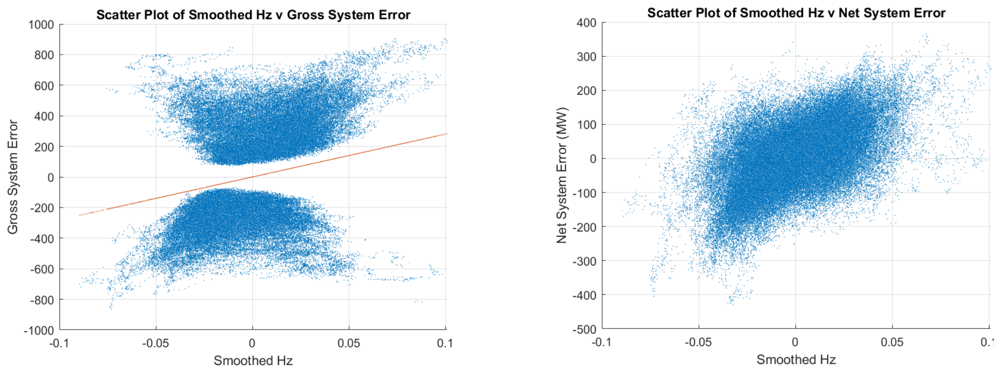
B.3.8 Gross or net system error as a combined metric

The concept of gross system error was outlined in Section B.2. This concept was explored as a possible metric on the basis that positive "good" deviations (those that are opposite in sign to the residual, or the majority) would tend to correct the errors in the system and "bad" deviations make them worse. Controlling deviations might be a more potent way to manage frequency than focussing on frequency. Possible advantages of this approach are:

- it could capture the performance of both PFR and regulation type responses, which are often difficult to separate in practice; and
- it could measure and remunerate PFR adequately by recognising and remunerating the work done under MPFR when there are greatly reducing frequency deviations.

To test this, we explored the relationship between gross system error as defined and a frequency metric. Smoothed frequency is one such measure. This is shown in Figure 19 following as a scatter plot from the first two weeks of September 2021. The charts show very little relationship between smoothed frequency and gross system error. While gross dispatch error under this approach is assumed to represent the population of good performers, not all of it is desired response like primary (droop) or secondary (AGC) action. Instead, gross system error includes "fortunate" dispatch errors of units that just so happen to be in the majority. At any one time with many units, and over many DIs for a single nit, this "netting" in the metered population was found to be substantial. On occasion it could significantly exceed the net error, particularly when the net error is close to zero, or when primary and secondary action can oppose each other, such as with time error correction or with a significant integral component in the regulation signal. Contrast the chart below with the neighbouring one of net system error, where the relationship is somewhat more evident but still not strong. Correcting for time error offsets appears to make little difference.

Figure 19: Scatter plots of smoothed frequency v gross and net system errors



•

The hope was that gross dispatch error would largely represent primary and secondary action, because the imposition of mandatory primary frequency response (MPFR) would increase population of metered elements under good control. This is untrue of the sample period where many inverter-based plant still had to implement the new droop settings.

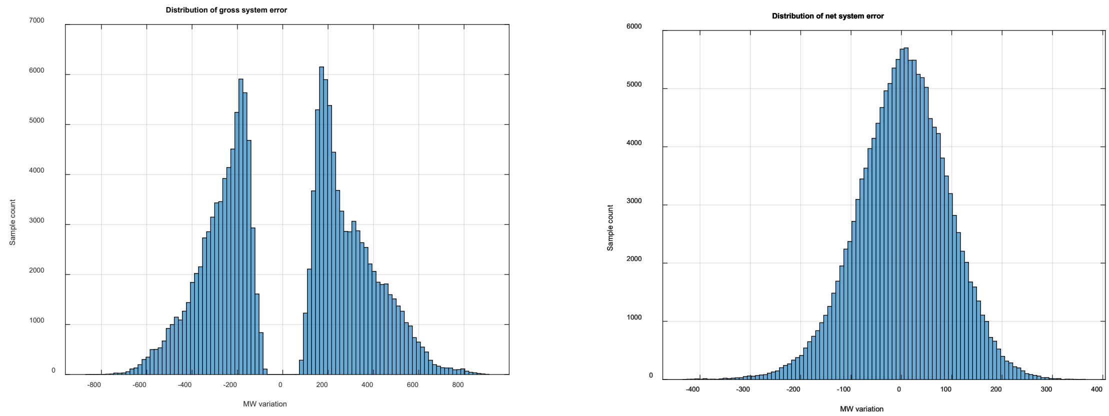
On a more positive note, it was evident the measure of gross dispatch error, by accounting for all deviations, inherently captured the mandated primary response. Given ubiquitous tight deadband MPFR means primary response is widely distributed over many units, large dispatch errors caused by only a few, or a significant forecast error can be absorbed on the power system without a sizeable frequency error, assuming secondary control adequately responds to the error to prevent the frequency error accumulating. The gross dispatch error provided a way of accounting for this error and was therefore of interest, not just as a performance metric, but also a scaling factor.

The project investigated the hypothesis that payments should be made of the order of the gross dispatch error, so the new arrangements would remunerate desired response like mandatory primary (droop) or secondary (AGC) action at the price of regulation FCAS. If the gross dispatch error also included "fortunate" dispatch errors of those units that just so happened to be in line with the gross dispatch, then these would also be paid, because to not do so might reduce payments for the desired primary and secondary response.

We were able to reproduce the broad shape of these measure using a simple simulation model. This work highlighted the role that random noise in the system is an important factor (noting that it also limited our ability to come up with a robust performance measure for individual

DIs). For example, the histograms in Figure 20. show how irreducible noise (in the short term) explains the hollowed-out centre of the gross system dispatch error distribution and the much wider spread of the gross system error relative to the net. As the system noise reduces, as it might well do with effective incentives and as more units implement mandatory PFR, these two distributions would come closer together.

Figure 20: Histograms of gross system error and net system error



Gross system dispatch error as a metric could be implemented in two ways:

- in its aggregate form, as a metric to be applied to individual unit deviations as a multiplicative performance measure; or
- as a means of classifying deviations as “good” (positive) or “bad” (negative) within a DI, the algebraic sum over the DI being the performance factor for that DI. This is equivalent to a +1 or -1 multiplier being used in the multiplicative formulation, depending on the “good” or “bad” classification.

An advantage of the second approach is that it recognises directly the work being done by PFR. A disadvantage is that the measure provides no guidance as to an appropriate response to the current frequency deviation, given the loose relationship between frequency and gross system dispatch error. This is a problem when the gross dispatch error may be high, but the net dispatch error is low – when it is hard to adjudge what is good performance. This may not be fatal if the focus of the arrangement is on the long run; to provide sufficient funds for investment in frequency performance capability. We will retain this option as a possible metric candidate.

To reiterate, at times of small net dispatch error with high netting in the metered population, there is a high gross dispatch error and it may be unclear what is good or bad – this suggests there is then a need to make deviations “less valuable” as compared to when there is a large net dispatch error. This is performed under the current Causer Pays system by multiplying the unit deviation by the performance metric – this feature should be retained. In this case it would be multiplying the unit deviation by the net system error.

B.3.9 Combined PFR and SFR frequency-based metrics

For a combined metric, it is easy to simply add or take a weighted mean of two metrics. For example, we could define:

$$Combined\_Metric = \frac{PFR\_\mathrm{Metric} + SFR\_Metric}{2} \tag{4}$$

Performance factor and therefore settlement calculations could proceed either separately or together, with identical results. So, for performance factor PF, we would have:

$$PF\_Combined(t) = \frac{PFR\_Metric(t) + SFR\_Metric(t)}{2} \times Pdev(u,t) = \frac{PF\_PFR(t)}{2} + \frac{PF\_SFR(t)}{2}$$

However, things are not so simple if we wish to split one or both metrics into Raise and Lower components. If raise and lower status is determined by the sign of the metric, the metrics when combined will cross the zero line at a time different to either of them separately. This may be appropriate, but it means that each component cannot be treated separately.

The AEMC has a policy position that the metrics should be combined, perhaps on the basis that PFR and SFR cannot easily be distinguished at any given moment. The crossover between Raise and Lower will be affected by this choice.

This may not be significant as crossover should typically occur when deviations are small and so where metrics and also performance measure values are also relatively small. Our default approach is to use a combined frequency-based metric, equally weighted, as this roughly approximates the share of the work done by primary and secondary control.

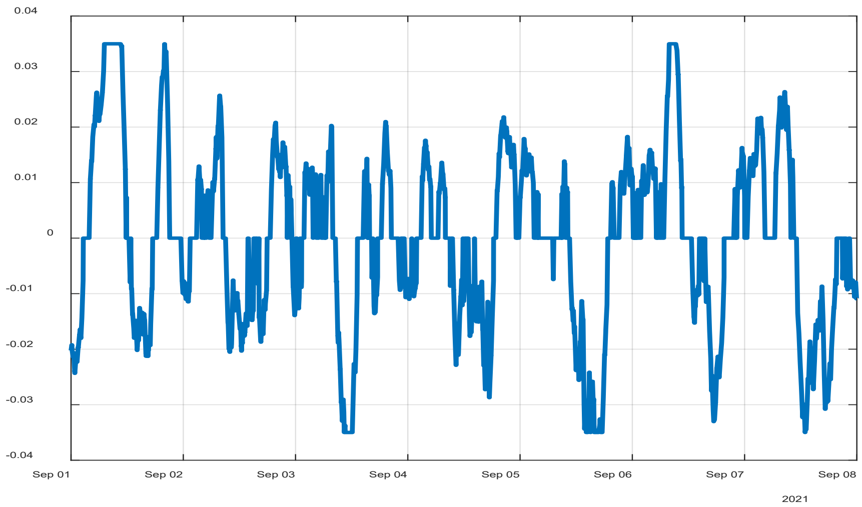
B.3.10 Time error correction

The AEMO AGC system has system for time correction designed to meet the time error requirements of the FOS. Once certain time error thresholds are reached, a frequency offset is delivered to the AGC for use as a target for unit control. Over time this frequency offset tends to correct any time error.

The offset has a lower (absolute) threshold and ramps up and down above that threshold in proportion to time error. The reason for this discontinuous approach probably relates to the dead band philosophy in the AGC, where little or no control is exerted if frequency and time error are within certain bounds.

Following is a plot of frequency offset (with sign reversed) over 1 week. Discernible is a daily pattern, albeit with a great deal of variability. What is striking is the bang-bang or on-off nature of the offset/control which leads to substantial offsets from time to time, well outside the maximum MPFR dead band. This in turn leads to periods of relatively high exposure to contingencies. Also, the offset-free proposed metric seems to work against what is being called for by AGC regulation at these times.

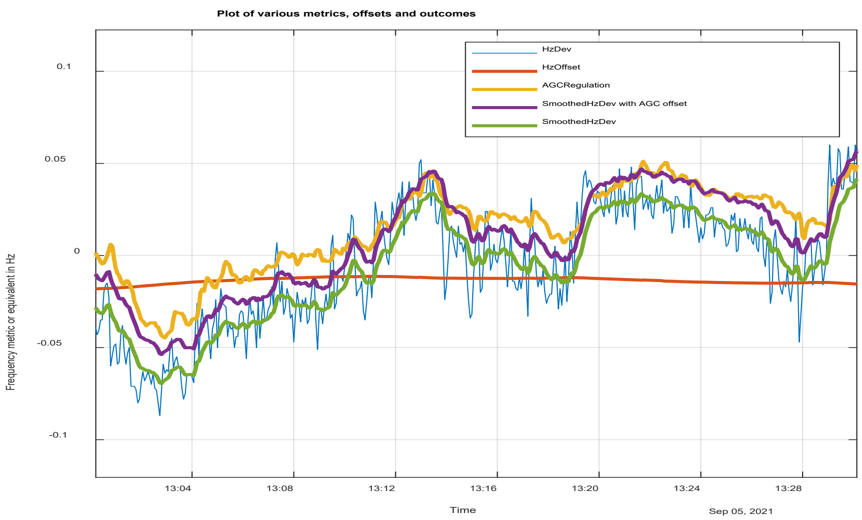
Figure 21: Profile showing offset correction from AEMO AGC



Source: AEMO 4 second data from September, 2021

Correcting for this offset makes a significant difference to the proposed secondary as shown in the following plot.

Figure 22: Profiles of potential metrics and system responses



Source: AEMO 4 second data from September, 2021

The chart shows the raw Hz metric in light blue and the smoothed Hz metric in green. Using the AEMO AGC bias of 2800, we can see that the AGC reg profile is well above. However, the AGC time correction offset is about 0.02 Hz negative over this interval as shown in red. If we apply this correction to the smoother Hz trajectory we get the purple trajectory which aligns much better with the yellow AGC reg output, as would be expected. Differences can be explained by the non-linear gains and the dynamic adjustments used in the AGC reg system for enabled units.

The AEMC has indicated a preferred policy position to omit any time error correction from the proposed incentive arrangements. This is driven by an understandable desire to avoid any direct connection with the AGC settings which may change in future.

However, even without a significant AGC offset signal a simple schema for a time error correction component in the settlement arrangement is possible and desirable. This can be achieved under the proposed performance-based incentive schema in two complementary ways.

- To the extent that the pressure that drives frequency offset exhibits a regular pattern, AEMO could define and publish a regular profile of frequency offsets, analogous to an energy market schedule. This profile would be always smooth and relatively small. It could be set to anticipate and minimise future expected departures from a zero-frequency offset.
- Next, we need a smooth, linear pricing signal to add or subtract an additional small offset to correct time error on the day. Consistent with the current proposed metric, this would be a slow-moving function of time error, converted to a frequency offset value. A possible metric component would be:

$$Time\ error\ metric\ component = Time\ error \times \frac{Base\ frequency}{Time\ constant} \tag{5}$$

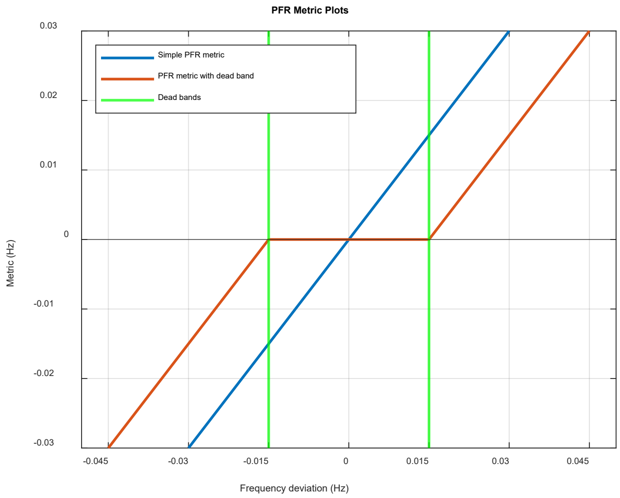
Time Error could be calculated as a simple integration of frequency deviation within the DI. This estimate could by synchronised with AEMO's time error periodically, say from data published periodically. Time error measurement drift is very slow so the need synchronisation is only occasional. The metric could be additive to the proposed frequency metric or the offset could be integrated into frequency metric. The former approach would be simpler to implement and manage in future

B.3.11 Reflecting the MPFR dead band in a PFR metric

The MPFR specifies a maximum dead band of 15 mHz on each side of zero. Participants may operate with a smaller dead band. The droop requirements do not apply from the zero point, but from the edges of the defined maximum dead band.

The issue is whether the metric intended to capture PFR performance should reflect the MPFR maximum dead band by having a dead band which reflects it. To illustrate, the figure below shows a simple, linear metric together with another that more closely tracks the MPFR maximum upper and lower dead bands.

Figure 23: Comparison of simple metric with one that follows MPFR dead bands



We have rejected this option for the following reasons:

- As a general rule, discontinuities in any incentives should be avoided as they may promote sudden and unhelpful responses (assuming participants pay attention at this level).
- The 15mHz band is a maximum and units may operate with smaller dead bands if it suits them. This should not be discouraged by removing the incentive at less than 15mHz deviation.
- Depending on the metric used, the value at risk within the dead band will typically be very small. In any case those not undertaking PFR inside 15 mHz will tend to have random deviations which will wash out over a billing week.

B.3.12 Reflecting an AGC dead band in an SFR metric

The AGC reg system implements a dead band structure, although its affect is somewhat disguised using a low pass filter to smooth out the raw signal. The question is whether an SFR metric ought to consider a similar dead band.

We argue against a dead band for an SFR metric for the same reason that we argue against one for PFR.

B.4 Conclusion

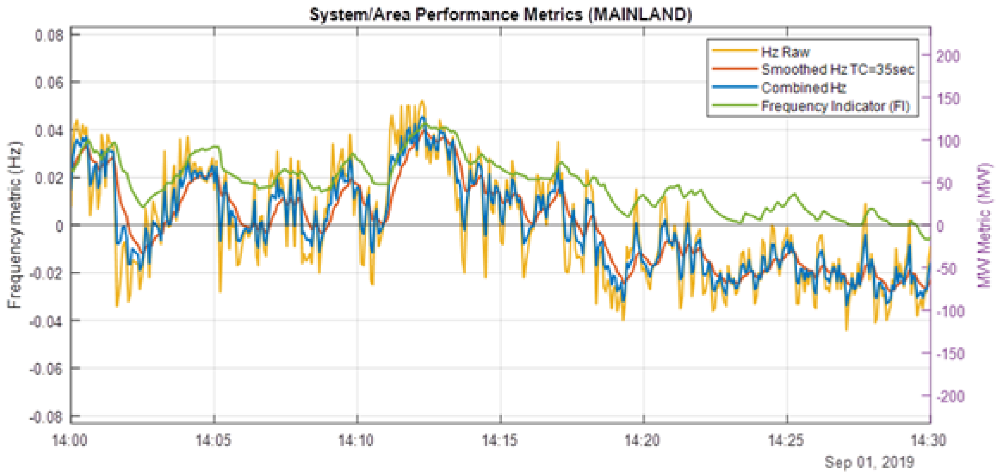
6.5.4. Summary of Investigation finding

- Measuring absolute performance with a measure such as correlation on a DI basis does not appear workable given the current pattern of frequency deviations.

- A gross system dispatch error approach could give a good indication of the potential scale of the errors to be corrected, although it appears unsuitable if used alone as a metric to manage frequency.
- A frequency metric targeting either PFR (raw Hz) or SFR (smoothed Hz) would be incomplete if implemented alone, as it would leave part of the requirement unserviced.
- The use of dead bands in any metric should be avoided.
- PFR, SFR and possibly other components can be managed combined or separated, but with slightly different outcomes if Raise and Lower are to be separated.
- The combined Hz metric proposed here takes no account of time error offset, which can be of the order of the offset observed.
- The use of dead bands in any metric should be avoided.
- PFR, SFR and possibly other components can be managed combined or separated, but with slightly different crossover points if Raise and Lower are to be separated.
- Frequency-based metrics for both PFR and SFR are the most direct ways to assess frequency control performance.

The figure below illustrates the frequency-based metrics considered, as well as the FI metric as a benchmark.

Figure 24: A sample of frequency-based metrics



AGC regulation has a relatively wide dead band which can inhibit it from crossing zero when frequency deviations are within the 15mHz or even greater. This behaviour may be active in the last third of the plot above.

It is impossible for a performance metric to perfectly reflect the control response, given that the power system is operated with primary and secondary controls that have different response times and lags. The performance metric need not be fully consistent with FI. A reasonable criterion is that a smoothed version or component of the metric should be the same sign whenever FI is outside the AGC regulation dead band. That is, AGC regulation and responses to the performance incentive should ideally not conflict, although they need not be of equal strength.

While initial investigations indicted that the combined Hz metric may best reflect the need for good frequency response, we chose four metrics to calculate performance factors and settlement results over a short (week period) period of real 4-second SCADA data. These were:

1. Frequency Indicator (FI) because it could act as a familiar benchmark for other metric options, despite its own drawbacks, which are:
    - it is an opaque metric generated by the AGC reg system which is very hard to reproduce locally;
    - it is currently recognised to need significant tuning; and
    - because it is likely to change in future, AEMC and AEMO policy is to avoid using it for the incentive arrangement
2. Smoothed Hz. With a time constant approximately matching the AGC, this metric could:
    - complement MPFR and AGC regulation to deliver improved frequency control
    - noting that MPFR responses would gain some benefits from this metric, even though it is not focussed on PFR.
3. Combined Hz, because it is a frequency metric that targets both PFR and SFR
4. Gross system error which, despite that it may not be a suitable metric because it poorly tracks frequency, is worthy of investigation because of its potentially good scaling properties.

For the gross dispatch error performance metric, we have shown that there can be times we significant unit deviations can occur in the metered population which substantially net off. Further, with tight deadband MPFR a high aggregate droop capability on the system may absorb a significant dispatch error without a large frequency deviation. For example, for the same dispatch error, the frequency deviation may differ depending on the aggregate droop capability.

A combined Hz performance metric may better identify good performance. If contribution factors are calculated as the product of unit deviation and the performance metric, occasions when there are significant deviations while the performance metric is insignificant (or ambiguous) are less important in determining financial settlement.

B.5 Reference Trajectory

B.5.1 Overview

Table 6-3: Base options and modifiers discussion - reference trajectory

Base option or modifier	Identifier	Discussion
Target-target trajectory	RT1.X.X	• Preferred option for most scenarios
Initial-target trajectory	RT2.X.X	• Rejected for reasons below • Will not value sustained responses across the dispatch interval
AGC Basepoint trajectory	RT3.X.X	• Rejected (Defer to David as well) • Basepoints do not reflect actual expected plant trajectory, but the setpoint that will get the plant to move along its expected trajectory.
No droop modifier	RTX.0.X	• Preferred option for all scenarios
With droop modifier	RTX.1.X	• Rejected as PFR would not be valued
No REG modifier	RTX.X.0	• See discussion below
With REG modifier	RTX.X.1	• Drawbacks of approach • an underlying (incorrect) assumption is that REG and PFR can be disaggregated • As a measurement dependant signal, it has the property of neglecting sustained PFR response • This will cause a hard link with AGC systems and this is not desirable as the AGC needs to be updated/replaced • This option was considered for one of the variations of scenario 3

B.5.2 Initial to target

The AEMO dispatch process takes the measured MW level at the start of a DI to determine the optimal target for each system unit. This target respects the unit's declared ramp rate limitations, capacity limit and bids. This initial to target straight line trajectory seems like a natural candidate for the reference trajectory.

As a unit will typically miss its target by varying degrees, this approach to the reference trajectory introduces a discontinuity in the MW value of the deviation. A unit deviation toward the end of a DI will lead to a financial exposure and a response if the metric suggests it. This exposure and potential incentive would seem to disappear when the trajectory is reset to the measured MW at the start of the next DI. This in turn could lower interest in controlling frequency at the start of a DI relative to the end.

In practice such an intuition may be pessimistic. An opportunity to respond at the end of the DI persists into the beginning pf the next DI even though the reference trajectory is reset; the incentive offered by the metric remains.

We put some weight on maintaining continuity in the deviations as measure across DI boundaries if that can be achieved with little practical penalty.

B.5.3 Target to target

An option that ensures continuity of deviation measurements is to define the reference trajectory as a linear ramp between adjacent dispatch targets. Relative to the initial to target option, this option nominally doubles the exposure of units by design. The upside is that we maintain a reasonably uniform incentive structure through the whole of the DI and reduce the risk of unsettling shocks at DI boundaries. The "downside" of higher exposure does not necessarily lead to any significant long run-run average change, as argued above.

A further question is: why a linear ramp? Answer: it's simple, transparent and consistent with what's expected of a unit, even though unit level implementation in the AGC may vary, as discussed in the following section.

B.5.4 Relevance of AGC base point

Some participants have argued strongly that the AGC base point is the only possible trajectory to use, because that is what units are being instructed by the AGC to follow.

The AGC base point signal can and often do differ significantly from the linear target-target trajectory? These departures can occur because:

- the linear schedule is not available until some tens of seconds after the start of the DI, possibly more allowing for communication delays; and
- units suffer various forms of lags in following their base points, so AEMO uses some unit-specific "nudges" to its signal in order to elicit an appropriate response, which is to reach the MW target as directly as possible.

Settling against the AGC base point may not be difficult administratively if ex post settlement is considered satisfactory. However, the AGC base point is *not* the trajectory that AEMO is seeking; it seeks a straight-line energy market trajectory modified by good response to frequency. The base point settings attempt to achieve that goal by compensating for transmission delays and sluggish unit response.

As noted earlier, deviation from the reference trajectory does not necessarily imply a penalty, the financial outcome depends on the frequency status of the system and the unit. Following the trajectory closely certainly reduces the variability of a performance payment but not much in terms of the average over a long period. Volatility is reduced further by the contribution factor approach, which measures a unit's performance against the performance of all others. Only a systematic correlation with the frequency-based performance metric, positive or negative, will elicit a significant ongoing payment, positive or negative.

B.5.5 Allowing for PFR in the reference trajectory

One option for the trajectory would be to add the MPFR maximum dead band to the chosen base energy market trajectory. In this way units under MPFR would be excluded from any exposure under a PFR incentive arrangement so long as they follow their scheduled energy trajectories and respond with PFR as required.

Generator Example

Energy schedule	340 MW
MPFR requirement	5 MW (frequency lagging)
Reference MW	345 MW
Measured Power	351 MW
Exposure	6 MW, to be settled under proposed incentive arrangement

This approach was considered and rejected for the following reasons:

- The required MPF performance is unit-specific, determined by negotiation with AEMO. It would require unit-specific settlement arrangements and would be difficult to administer.
- The approach would neutralise any payment due to the provider for compliant PFR performance. This is not consistent with the intent of the current rule change.

B.5.6 Including AGC regulation in the reference trajectory

The intuitive logic for including AGC regulation in the reference trajectory is that it would:

- reflect the trajectory that units on AGC regulation are expected to follow; and
- avoid 'double payment' for AGC regulation

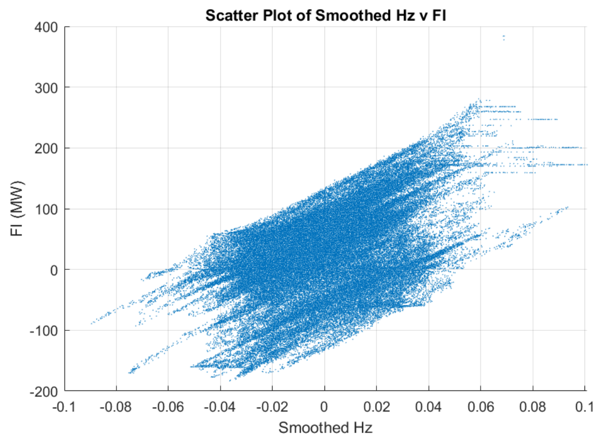
There are some potential implementation issues to be kept in mind:

- Use of the real time AGC trajectory signal as a component of the reference trajectory violates the assumptions behind the top-down approach to calculating the residual – see section B.8
- The AGC signal is a function of the measurement of the unit as well as the frequency; hence it does not form a solid reference against which to measure power and performance.
- Any scaling based on gross error or equivalent should be adjusted down to reflect the work being performed by AGC regulation, which would be effectively removed from the proposed performance incentive system.

If the AGC regulation trajectory is excluded from the reference trajectory, units under AGC regulation would be required to follow AGC regulation control signals that are not fully aligned with the performance metric to which they will be substantially exposed. The issue is whether this exposure would be advantageous or detrimental to such units and, if so, potentially detrimental to good frequency control.

The figure below shows a plot of smoothed frequency (the proposed SFR metric) and FI, the metric that drives the total response from units under AGC regulation. While imperfect, there is a clear positive relationship between the two metrics. This is not surprising as the metrics are both based on frequency and similarly smoothed. Differences are that FI has a variable gain (and a significant dead band in particular), time error correction and dynamic adjustment within the DI. The key point is that an AGC regulation-enabled unit that is performing well can expect to benefit on average from the incentive arrangement, even though there may be times when the two measures are not fully aligned. Bad AGC regulation performers would have no such assurance, of course.

Figure 25: Relationship between FI and smoothed frequency with TC = 35 secs



As the incentive arrangement would apply to all units, to a first order the relative attraction of being enabled or not would not change. However, some units currently working under AGC regulation may feel they would do better outside if they could operate more profitably under a performance incentive only. Other units may opt to become AGC regulation enabled if they find the control offered by the AGC regulation to be convenient. Either development could be considered a benefit to the system.

B.5.7 Conclusion

From the discussion ion this section we conclude that:

- Initial to target and target to target straight line trajectories are both viable reference trajectories.
- To always ensure consistency of incentive, the target-to-target trajectory is preferable. The possibility of slightly higher exposure would not lead to a significant change in long-run financial outcomes.
- Including the AGC reg trajectories has the effect of greatly reducing AGC reg units' exposure to most performance payments, but if poorly implemented could measure these units more harshly than non-enabled units.

- Including the AGC trajectory on the reference trajectory is not recommended as it potentially distorts the relationship between enabled and non-enabled units.

B.6 Classification

B.6.1 Overview

Table 6-4: Base options and modifiers discussion - classification

Base option or modifier	Identifier	Discussion
Positive/Negative performance before 5min aggregation	C1.X.X	• Preferred option as 4-second incentives will be maintained. There will be no opportunity to net out performance.
Positive/Negative performance after 5min aggregation	C2.X.X	• See above
No Raise/Lower classification	CX.0.X	• See below
With Raise/Lower classification	CX.1.X	• Preferred to reflect different cost functions for providing raise/lower services, which in turn warrant separate financial weightings applied to different directions • This was used in all scenarios

B.6.2 Determine performance before or after aggregation

Possible approaches to measuring performance are:

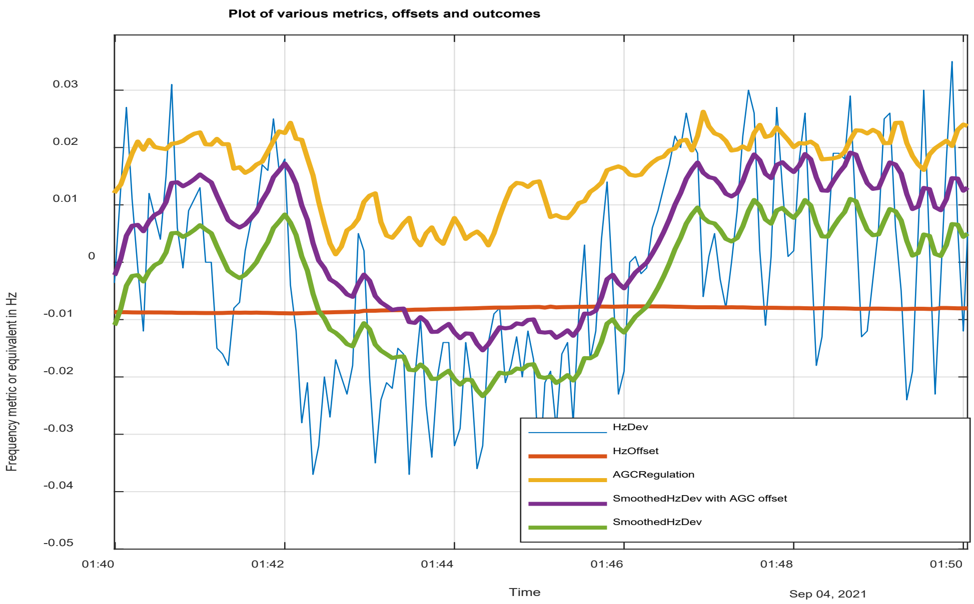
- Apply the performance metric at each measurement point, and then aggregate to the DI level for financial weighting at the DI level.
- Aggregate net deviations within a DI and apply a single performance metric at the DI level.

A similar choice is possible when deciding what factors go into the Raise and Lower Categories as discussed below. We could choose to maintain Raise and Lower buckets at the 4-second level or make that decision only at the 5-minute level, on some form of average basis.

The first approach is preferred, for two reasons.

- A typical plot of frequency behaviour (for example, in Figure 26.) shows potential wide variation within a DI. If performance is only measured at the DI level, all this detail is lost and the effectiveness of the metric reduced.
- a robust metric should not depend on the chosen level of aggregation used for administrative convenience– in this case 5 minutes.

Figure 26: Typical variability of frequency and related measures over two DIs



We note this this is a significant distinction at the DI level, although it may become less significant over a billing week as the netting out process over that period would likely yield a similar settlement result.

B.6.3 Separation of Raise and Lower

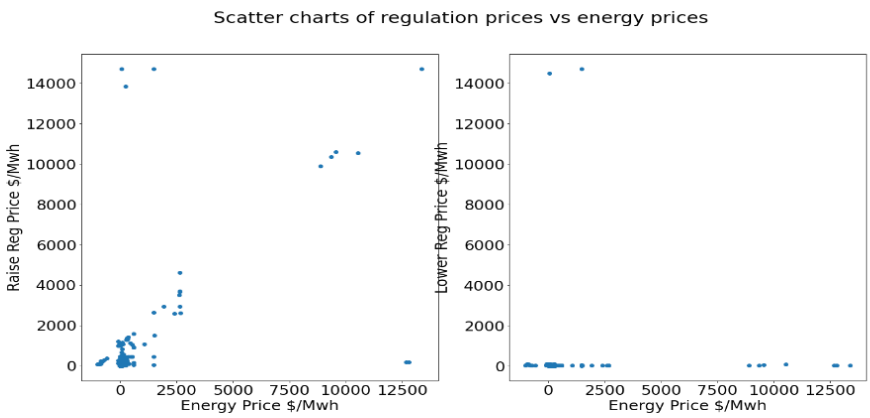
For SFR, separation of Raise and Lower seems like a natural extension of the approach taken with centralised AGC regulation. The justification for a split lies in the much lower cost of Lower relative to Raise, as illustrated in the plots of Figure 27 following. The Raise plot also shows some relationship between energy price and raise prices, especially at high energy prices, but not so for Lower. To implement a split between Raise and Lower, the system would need to:

- separate the performance measurements at the 4 second level according to the sign of the metric (positive Raise, negative for Lower); and
- apply a different financial weight to each of the Raise and Lower buckets at the DI level.

Separating Raise and Lower would complicate implementation, reporting and management of the incentive arrangements slightly, but not unduly. In FDP terms the approach would present a discontinuity in the slope of the price curve at the negative/positive boundary. Again, this would be manageable. Separation of Raise and Lower is already performed in the current Causer Pays arrangement, to separately account for the different costs of Raise and Lower.

Experience in the longer term may demonstrated that separating Raise and Lower may be an unnecessary complication for a service that is not scheduled. Therefore, we will recommend that the matter be revised after a period of, say, two years' experience with the incentive.

Figure 27: Relationship between energy and regulation prices (in MWh) – Raise and Lower



B.6.4 Conclusions

- Positive or negative status of performance factor should be determined at the point of measurement rather than the DI level
- Separating Raise and Lower is the preferred mode for an initial implementation.
    - However, this should be reviewed after two years' experience
    - Raise and Lower status should be determined at the measurement level.

B.7 Aggregation

B.7.1 Overview

Table 6-5: Base options and modifiers discussion - aggregation

Base option or modifier	Identifier	Discussion
Settlement by unit	UA1.X	• Preferred option for simplicity because it: ○ avoids the shortcomings in the current system ○ Is not relevant in any case for the balanced performance incentive component
Settlement by portfolio	UA2.X	• This was not investigated in detail
Sum/Mean of performance functions	UAX.1	• This was considered in a variation of scenario 3. • This variation makes no substantive difference to outcome but supports a simpler and more robust implementation.

Base option or modifier	Identifier	Discussion
sum of metered performance factors over **gross** system performance (for unit-level performance factors)	UAX.2	• This was considered in most scenarios (except scenario 3.3)
Mean of performance factors over **net** system performance (for unit level performance factors)	UAX.3	• This is relevant only to scenario 3.4 (uses the relative performance scaling option). • The mean is chosen over the sum as the scaling option uses short-term vs long-term performance of net-error and both short term and long-term performance needs to be on the same scale (either 4-second or 5-minutes).

B.7.2 Settlement by Unit or Portfolio

As a matter of principle there should be no advantage offered to portfolios by settling at the portfolio level. Therefore, all settlement should take place at the unit level for scheduled units. Settlement at the portfolio level was not investigated. In any case it is expected to offer no advantage under the current proposals for performance payments.

For cost allocation of AGC regulation enablement, the cost allocation is one sided and some advantage could be gained by a portfolio from settlement at the portfolio level. However, we do recommend against it for this case also, for the same reason.

B.7.3 Use of means or totals for performance factor

The performance factor over 5-minute could be expressed first, as a simple sum, as in:

$$PF(u,di) = \sum_{t \in di} PF(t) \tag{6}$$

or, alternatively

$$PFmean(u,di) = \frac{\sum_{t \in di} PF(t)}{\text{nsamples}} \tag{7}$$

where t is a time of measurement, $nsamples$ is the number of sample points in the DI (75 in the case of 4-second data and 5-minute DIs). This choice is of no consequence in principle as PF either occurs ratio of single and aggregated factors to form a contribution factor, or the term $nsamples$ can be accounted for in the scaling factor. However, the use of the mean form has advantages in implementation. Specifically, it is independent of the measurement interval, thereby avoiding the need for interpolation (in the case of measurement intervals greater than 4 seconds) or losing available accuracy by averaging.

B.7.4 The denominator in contribution factors

The Draft Rule proposed will retain a contribution factor approach, whereby each unit would be assessed relative the whole, to determine a share of a total amount paid in the DI, however that is determined. As the project progressed, we noted that there were two possible approaches to determining this share:

- The net approach, where the DI-level performance factors of all scheduled units are aggregated/netted, i.e. no separation of positive or negative factors at 4-sec level. These deviations will be of different signs (equivalent to C2.X.X in section 3.5). The denominator is the negative of this sum, or the Residual deviation, according to the logic for determining the Residual described in B.8. This ratio is guaranteed to be one, which would share out all of the determined dollar amount.

- The gross approach, where only the positive (or negative) DI-level performance factors are aggregated. The denominator is the absolute sum of the positive (or negative factors). Positive and negative factors sum to zero by definition[9].

These approaches achieve different ends. The net approach would set an amount that would be allocated to the Residual, with the same amount (but opposite in sign) being allocated to scheduled units. Some of these units would have negative performance factors, so the amount allocated would be the net of the positives and negatives.

The gross approach would target the turnover i.e. the total payments between providers and causers – between the positive factors and negative factors.

AEMC policy is to avoid rules that specifically target currently unscheduled units because, in future, under ESB long-term thinking, the market may become two sided and fully scheduled. In any case, the proposed incentive arrangements would be focussed on the total requirement for frequency control, which is substantially a gross requirement. Aggregate gross performance factors and associated settlement amounts are also much easier to interpret than net amounts. Use of gross aggregation is our preferred approach

B.7.5 Conclusion

- The incentive system should distinguish Raise and Lower in its initial implementation with a suggested review to see where a single combined service could be viable and simpler after two years of operation.

- For implementation, consider using the mean of performance factors rather than the total, as this makes processing independent of the measurement interval.

- Gross rather than net amounts should be used in settlement formulae where appropriate, reflecting the combined requirement of PFR and SFR.

[9] Including the AGC trajectory in the reference trajectory might require this assertion and the conclusions that flow from it to be adjusted.

B.8 Residual Calculation

B.8.1 Overview

Table 6-6: Base options and modifiers discussion - residual calculation

Base option or modifier	Identifier	Discussion
Top-down	RC1	• Preferred for simplicity and will result in balanced performance factors which will result in balanced performance payments
Bottom-up	RC2	• See above

6.5.5. Concepts

All scheduled units are required to have SCADA-level metering installed for monitoring and control. The frequency control incentives Rule would allow use of such metering to settle the payments due. Load and embedded generation not metered in this way will also participate indirectly by changing load passively and help fully as frequency changes. Also, AEMO makes demand forecasts on behalf of this unmetered (by SCADA) group. Forecast error is a large component of the need for frequency control.

This ‘unmetered’ demand is called the residual. The residual consists not only non-metered load and generation but also real power losses in the system. Viewed in this way we see that the residual can be measured. We just sum all the meter readings in the system and change the sign. This is valid because the sum of input and outputs to and from a set of meters must always sum to zero.

For settlement we need the deviations of the residual at each measurement (SCADA) time. There are two broad approaches

6.5.6. Bottom Up

This the approach implemented in the current causer pays system. The residual is broken up into components.

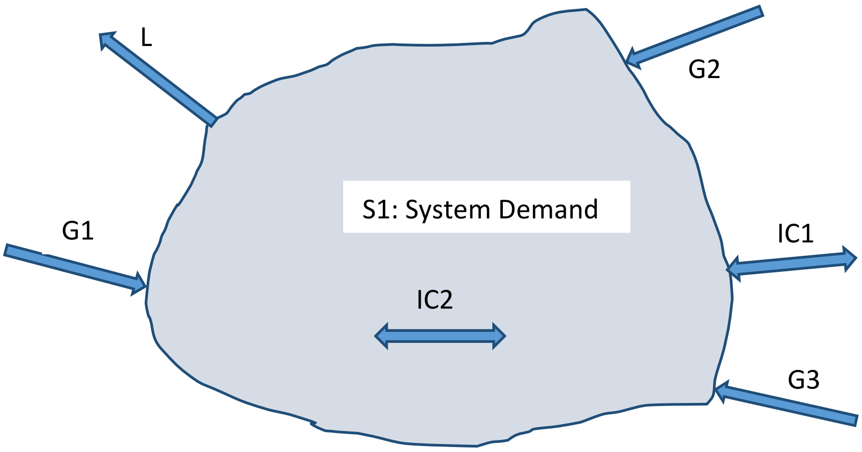
- A straight-line forecast error is first determined as the difference between the forecast made ex ante and used for scheduling and the end point of a straight line of best fit using metered data.
- Aggregate meter deviations from the straight line are then calculated.
- Separate assessments are made of interconnector losses between regions, as these are not accounted for in the steps above.
- Because there will always be errors in this approach so that inputs and outputs do not match, the resulting input and outputs must be scaled so that they balance physically and, ultimately, financially well.

This procedure is complex and non-transparent. It may well be useful to distinguish these components for management purposes but they add nothing to settlement.

6.5.7. Top Down

The figure below shows a simple system with a boundary drawn around all the entry, exit and combined metering points. This boundary can be drawn around any set of meters of scheduled units that fully enclose the whole or part of a part of a synchronous system. This example includes an internal DC link IC2.

Figure 28: Illustration of inputs and outputs to a settlement region



- At the time of scheduling, AEMO makes a forecast, SI, and all other units are scheduled so that the Residual is given by:

$$R = -\big(G1 + G2 + G3 + IC1(out) + IC2(out) - L1 - IC1(in) - IC2(in)\big) \qquad (8)$$

  This balance is enforced in the dispatch schedule produced energy dispatch and pricing engine, albeit with modelling approximations. The negative sign ensures that all these elements, including the residual, sum to zero.

- After the event, the above equation must also apply, by the laws of physics, for the physical metered units, subject only to metering errors. This relationship is another way to confirm the logic discussed above of summing the meter readings and reversing the sign to derive the effective residual meter readings.

- We seek now to calculate the Residual deviation for settlement purposes i.e. we want.

$$R_{deviation} = R_{metered} - R_{reference} \qquad (9)$$

If we apply this definition to the balance equation above, it is easy to see that the residual deviation is the negative sum of all the metered deviations.

This is a useful result because it guarantees not only a physical balance but also a balance of settlement payments. However, various additions, subtractions and normalisations applied later may complicate this result.

B.8.2 Conclusion

- On the grounds of robustness, simplicity and transparency, we propose to implement the top-down approach to determining residual deviations.
- A more detailed breakdown of contributions to residual deviations for management purposes can be performed in processes separated from settlement.

B.9 Financial Weight

B.9.1 Overview

Table 6-7: Base options and modifiers discussion - financial weight

Base option or modifier	Identifier	Discussion
Separate regulation costs	F1	• Assumes separate Raise and Lower – Draft Rule settlement formula
Separate regulation prices	F2	• Assumes separate Raise and Lower – Proposed settlement formula
Energy price	F3	• Could be an element of the weight if the service is combined, but need to adjust for very low and negative energy prices
Combined regulation cost	F4	• Not considered
Combined regulation price	F5	• Option for a combined service

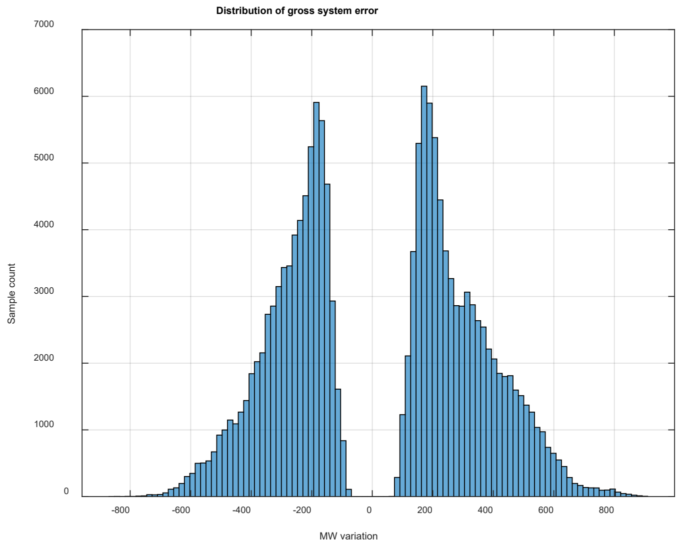
Choice of financial weight is substantially driven by a decision on whether to combine or separate Raise and Lower services.

B.9.2 Separate Raise and Lower services

This sub-section covers options F1 and F2 on the assumption that Raise and Lower are to be separate services and separately weighted.

The Draft Rule would have implemented option F1. In each DI, for each of Raise and Lower separately, the total cost of (Raise of Lower) regulation, labelled *TSFCAS* in the Draft Rule, would be assigned as a total performance payment, scaled up or down depending on the volume relationship between the Regulation Requirement (RR) and Enabled Amount (EA). It would be allocated according to contribution factors, or shares, based on measured performance.

Figure 29: Histogram of Gross System Error



On review, this approach was rejected. Whilst mathematically identical, it was considered better to simply use the enablement price for weighting and then to derive a separate scaling factor or volume, against which the enablement price could be scaled. If gross system error is used for scaling, the histogram above would indicate a scaling several times larger than the 200 MW typically enabled for AGC regulation. This is option F2.

There are other possibilities for setting the financial weight; for example, by setting a price based on the supply curve revealed in the enablement market to separately price PFR. The argument is that PFR is more valuable than regulation and should be priced accordingly. On the other hand, there is some evidence from the data that PFR is being used substantially ahead of regulation, so discounting SFR relative to PFR seems inappropriate in that situation. It is impractical to reach a conclusion on this issue in the absence of practical experience under a proposed arrangement. This highlights the case providing for a review of the system after, say, two years of operation.

B.9.3 Combined Service

This covers options F3, F4 and F5

For the reasons outlined above, we did not consider the cost-based approach of F4.

One way to preserve the difference in Raise and Lower to some degree in a combined service is to use:

$$Financial\ Weight = \max(RegRaisePrice, Reg\text{Lower}Price) \tag{10}$$

Or we could cover all bases with:

$$Financial\ Weight = \max(EnergyPrice, RegRaisePrice, Reg\text{Lower}Price) \tag{11}$$

This was not examined in any further detail as the preferred approach is to maintain separate Raise and Lower Services

B.9.4 Conclusion

- Assuming separate Raise and Lower Services the preferred approach is option F2, use regulation enablement price as the financial weight.
- In the combined service case, there are several possibilities involving energy price and regulation prices. These have not been examined in detail as a combined service is not the preferred approach.
- There is a case to review the pricing arrangement and other matters after a period of operation of, say, two years under the new incentive arrangement.

B.10 Scaling

B.10.1 Overview

Table 6-8: Base options and modifiers discussion - scaling

Base option or modifier	Identifier	Discussion
Relative need	S1	• This was the Draft Rule approach – RR/EA • The RR term is the maximum of some MW measure of the requirement, based on net or gross concepts. • Benefits: o If the numerator reflects total response (not net response) then this will reflect the value/size of non-enabled good response • This is considered in nearly all scenarios (except 3.3 and 3.4).

Operator's constant	S3	• Benefits
		  o Provides AEMO discretion in setting the total turnover to meet system objectives
		• Drawback
		  o As a matter of principle, giving AEMO the discretion to directly influence settlement outcomes is not a good idea
		• This option is considered in scenario 3.3

Relative performance	S4	• Scales amounts by short-term performance and long-term performance of the net error. Payments are scaled up if short-term performance exceeds long-term performance. An extra scaling parameter may be required to adjust the turnover (as in S3). • This is considered in scenario 3.4

B.10.2 Analysis performed

These options were arrived at after considering the outcomes from base scenario runs (refer to Scenarios 1–4 in the short sample period results in Section 5. These base Scenario runs showed quite small turnover from the performance incentive, bar the option that used gross dispatch error. Such a small turnover, commented on by participants who did their own calculations, could be insufficient to justify investment in frequency control capability.

On analysis of these results, the team judged the reasons for the low turnover to be:

- the implied use of net MW figure as the MW scaling factor RR/EA in the Draft Determination. This AGC regulation quantity was deemed insufficient to cover off the requirement for both PFR and decentralised regulation. The requirement was judged to be based more on the "above" or gross dispatch error, which was typically much larger than the AGC enabled amount; and
- the mediocre performance of units assessed against a metric not yet implemented.

B.10.3 Relative Need

This option is an implementation of the Draft Rule scaling RR/EA – hence the label "relative need". RR, regulation requirement is an undefined term which generates a range of specific approaches. EA is the AGC regulation enablement amount.

Our initial runs implemented this concept verbatim. RR was initially taken to be the maximum of the AGC regulation total requirement (as measure by FI) over a DI. The team realised that the draft rule could be simplified:

$$FPP = Contribution\_Factor \times Regultion\_Cost \times \frac{RR}{EA}$$

Is equivalent to:

$$FPP = Contribution\_Factor \times (Regultion\_Price \times EA) \times \frac{RR}{EA}$$

Which simplifies to

$$FPP = Contribution\_Factor \times Regultion\_Price \times RR$$

This simplification clearly displays the term RR as a scaling factor, usually in MW bur possibly in some other form if the contribution factor is not normalised

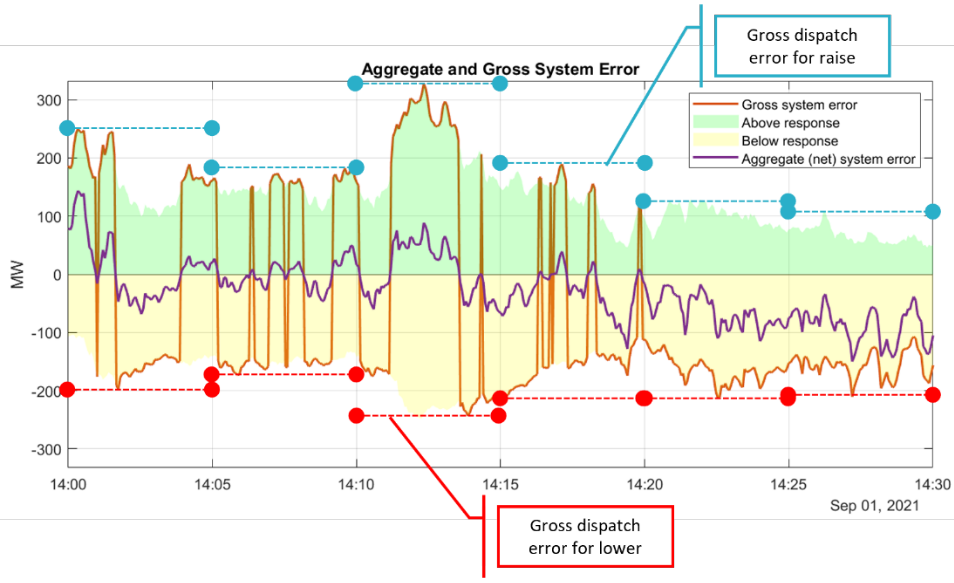
The scenario analyses discussed in Sections 5 and 6 give an extensive discussion of hoe arrived at our preferred options for scaling. These are two approaches to a gross system error scaling.

Gross system Error

Our investigations pointed to a gross system error giving appropriate scaling, at least at the macro (total settlement) level. The argument is that gross system error is a measure of all the deviations that need to be covered. Although at present many of these deviations are system noise, such deviations will tend to net out at any moment across many units, or for a single unit over time.

Gross dispatch error scaling was implemented as the maximum of the gross system error over a DI, as illustrated the example below.

Figure 30: Gross dispatch Error scaling



Inverse Hz Scaling

This logic arises from an FDP line of analysis. The approach is considered in some detail in APPENDIX C.

The basic FDP formula is:

$$\frac{FDP}{Reference\_Price} = \frac{Frequency\_Deviation}{Reference\_Frequency\_Deviation}$$

If $Reference\_Frequency\_Deviation$ Is set at some threshold, say the MPFR dead band of 15mHz, this defies the point when the FDP equals the Reference Price, which is the Regulation Price . This in turn defines a scaling which appears as a $\frac{1}{Reference\_Frequency\_Deviation}$ term in the FDP settlement formula.

B.10.4 Operator's constant

This approach recognised that the positive and negative payments in the performance incentive system (distinguished from the AGC enablement cost recovery) are naturally balanced within a settlement region, due to the physical energy balance (including losses) of real power within the boundaries of a system. Therefore, it is technically easy to apply a simple multiplier to the basic settlement formula to get a valid settlement regime with a different turnover; in fact, any desired turnover.

However, this approach is considered to allow AEMO far too much discretion to affect settlement outcomes. This approach was therefore rejected, although we will recommend that a strategy to "scale up the scaling factor' of the settlement formula to a long-term level over the initial 12 months of the arrangement would have merit, to allow participants and AEMO to tune their systems and strategies progressively.

B.10.5 Relative performance

This approach arises naturally from an FDP approach to settlement assuming a net normalisation factor as outlined in Section B.7.3. As we have ejected the use of a net measure, we have rejected this scaling option also.

However, an FDP approach is still viable without normalisation. This has been considered as an option in the Relative Need option in Section B.10.3 above

B.10.6 Conclusion

A good scaling approach appears to be a critical element of the procedure. Option S1 together with an FDP variation, or come combination of these, seem most promising. These options are examined in more detail in the scenario studies described in Sections 5 and 6. A deeper investigation of outcomes at the DI and unit level under different situations such as high and low requirements is also recommended to support implementation choices.

B.11 Enablement Cost Distribution

B.11.1 Overview

Table 6-9: Base options and modifiers discussion - enablement cost distribution

Base option or modifier	Identifier	Discussion
Used cost from poor performers + unused from all units	EC1	• This is the Draft Rule approach.
All cost from poor performers	EC2	• Not considered
All cost from non-metered	EC3	• This is used in scenarios 3.3, 3.4 and 4
Used cost from poor performers, unused cost from non-metered	EC4	• This is used in all scenarios (except 3.3, 3.4 and 4)
Used cost from poor performers, unused cost accorded to some average of prior performance	EC5	• Used in final scenario runs

B.11.2 Analysis

Option EC1 – used costs to poor performers; unused costs from all units

This approach from the Draft Rule was criticised by participants who argued that allocating unused costs to scheduled unit would distort their behaviour. Allocation to loads (i.e. the residual) would be less distortionary because loads generally do not know or cannot respond to sch a charge.

This criticism was of sufficient force to seek an alternative, less distortionary approach.

Option EC2 – all costs from poor performers

This would be an undue burden on poor performers and not considered further

Option EC3 – all costs from non-metered (residual)

This option was used in some scenarios (Scenarios 3.3, 3.4 and 4) in early (two week) settlement runs. However, AEMC policy is to avoid allocating costs to the residual as loads may in future become active market participants under ESB long-term proposals. This approach was abandoned in later scenario runs in favour of option EC5.

Option EC4 – used cost from poor performers, unused cost from non-metered

Allocating costs in this way was a less distortionary compromise that was used in all early scenario runs other the 3.3, 3.4 and 4. However, it was abandoned in favour of option EC5 based on the AEMC's policy position.

Option EC5 – used cost to poor performers unused cost from prior performance

This option is an attempt to reduce the potentially perverse incentives from allocating unused enablement costs while avoiding allocating these costs to the residual. While the final details are still to be worked out, smearing these costs based on a sequence of historical performance factors minimises the risk of bad behaviour when costs are high.

B.11.3 Conclusion

Given the AEMC policy position is to avoid allocating costs to loads through the residual, option EC5 is a workable compromise.

APPENDIX C. Analysis of an FDP Approach to Scaling

C.1 Introduction

This appendix uses a FDP approach to derive an FDP settlement formula consistent with the gross system error scaling approach.

C.2 Analysis

The FDP approach begins with the pricing of instantaneous energy according to a formula structured as follows.

$$\frac{FDP}{Ref\_Price} = \frac{fdev}{Ref\_fdev} \times Constant \tag{12}$$

Where

FDP is Financial Deviation Price

Ref Price is a suitable reference price (e.g. regulation price)

fdev is frequency deviation or, more generally, any frequency-based metric

Ref fdev A reference frequency

Constant A dimensionless constant

Re-arrangement emphasises the direct proportionality between the *FDP* and *fdev* within a DI because the last term contains values that are constant and known in advance of the DI.

$$FDP = Ref\_Price \times \frac{fdev \times Constant}{Ref\_fdev} \tag{13}$$

fdev is a proxy for the weighted sum of the raw and smoothed frequency values of our chosen metric. We do not need to make this assumption, but it simplifies the algebra that follows.

This basic form is justifiable in control theory but also in the analysis that follows. Note that *FDP* and *Ref Price* are both in units of \$/MWh, so that *Constant* is dimensionless.

From this we can derive a settlement amount for a unit within the DI as the sum of small energy increments over the measurement interval *delt.*

$$FPP(u, di) = \sum_{t \in di} FDP(t) \times Pdev(u, t) \times delt \tag{14}$$

where

FPP is Frequency Performance Payment

Pdev is the MW deviation of unit u at measurement time t

delt is the interval between measurements

Substituting FDP from (13), adding suitable arguments and re-arranging we get, we get:

$$FPP(u, di) = \left(\sum_{t \in di} fdev(t) \times Pdev(u, t) \right) \times delt \times Ref\_Price(di) \times \frac{Constant}{Ref\_fdev} \quad (15)$$

For convenience, define a Mean Performance Factor $PFmean(u, di)$ as follows:

$$PFmean(u, di) = \frac{\sum_{t \in di} (fdev(t) \times Pdev(u, t))}{nsamp(di)} \quad (16)$$

Plugging this definition into (15), we get:

$$FPP(u, di) = PFmean(u, di) \times nsamp(di) \times delt \times RefPrice(di) \times \frac{Constant}{Ref\ fdev} \quad (17)$$

We note that *nsamp(di) x delt* is equal to the duration of the DI and that *Ref_Price* is in $/MWh.

Plugging this into (17) we get:

$$FPP(u, di) = PFmean(u, di) \times Ref\_Price(di) \times duration(di) \times \frac{Constant}{Ref\ fdev} \quad (18)$$

This has a form along the lines of that defined in the Draft Rule as proposed to be amended, namely:

- a performance measure;
- a financial weight
- an additional scaled quantity.

However, these components differ in detail from other approaches considered in this work. Specifically, the MW volume relating to the gross error in the system which is to be corrected is not immediately evident in this formula. We need to relate this approach more closely to the standard form and to determine a justifiable value for the dimensionless scalar *Constant.*

The first step is to sum *FPP(u,di)* in (18) over all units with positive performance factors. Noting that Raise and Lower are separately treated, we can do this by summing over all units with positive deviations for Raise and all units with negative deviations for Lower. We retain the same *nsamp* count (*nsmap* is 75 for 4-second metering and 5-minute settlement) for both Raise and Lower. In this way the two Raise and Lower services can simply be added, albeit with different financial weights.

To indicate this gross summation, we simply remove the "u" argument in (18) to get, for either Raise or Lower:

$$FPP(di) = PFmean(di) \times Ref\_Price(di) \times duration(di) \times \frac{Constant}{Ref\_fdev} \quad (19)$$

We can do this for *PFmean* because our metric is common to all units. We now seek to bring this formula into closer alignment with the standard AEMC approach by dividing and multiplying by a l*ong-term* mean Performance Factor, *PFmean*:

$$PFmean(lt) = \frac{\sum_{t \in lt} fdev(t) \times Pdev(t)}{nsamp(di)} \tag{20}$$

This long-term measure demonstrably stabilises over a relatively long period such as 4 weeks. DI level performance factors can be greater or less than this long-term value.

Multiplying and dividing by (18) by this value, we get:

$$FPP(u,di) = \frac{PFmean(u,di)}{PFmean(lt)} \times Ref\_Price(di) \times duration(di) \times PFmean(lt) \times \frac{Constant}{Ref\ fdev} \tag{21}$$

Or:

$$FPP(u,di) = CFLT(u,di) \times Ref\_Price(di) \times duration(di) \times PFmean(lt) \times \frac{Constant}{Ref\_fdev} \tag{22}$$

Where *CFLT* is a DI level contribution factor relative to the long-term. These factors will sum to less than or greater than 1, depending on what happens within the DI, but will approach 1 on average over many DIs. With this small but important difference, it resembles a within-DI contribution factor.

Now let us examine the last two terms in (22) in more detail i.e. the terms:

$$PFmean(lt) \times \frac{Constant}{Ref\_fdev} \tag{23}$$

We first note a close relationship between *PFmean(lt)* and the covariance between frequency and gross dispatch error $Pgrossdev$ over the same long-term period. By definition:

$$cov(fdev, Pgrossdev) = \frac{\sum_{t \in lt}(\, fdev(t) \times Pgrossdev(t)\,)}{nsamp(lt)} \\ -mean(fdev) \times mean(Pgrossdev) \tag{24}$$

Noting also the definition of *PFmean* in (20) we therefore have:

$$cov(fdev, Pgrossdev) = \ PFmean(lt) - mean(fdev) \times mean(Pgrossdev) \tag{25}$$

We can be assured that *mean(fdev)* approaches zero as the sample size becomes large because the time error is capped as a matter of operational policy. It follows that, to a high level of approximation:

$$PFmean(lt) = cov(fdev, Pgrossdev) \tag{26}$$

From the definition (24) and simplification (25) applied to (6), we can now write:

$$PFmean(lt) = cor(fdev, Pgrossdev) \times std(fdev) \times std(Pgrossdev) \tag{27}$$

The advantage of this re-formulation is that correlation is a well understood and easy to calculate with standard functions.

Returning now to the expression in (23) we can plug into (27) to get

$$Pfmean(lt) \times \frac{Constant}{Ref\ fdev} =$$

$$cor(fdev, Pgrossdev) \times std(fdev) \times std(Pgrossdev) \times \frac{Constant}{Ref\_fdev} \quad (28)$$

Setting *Ref_fdev* to *std(fdev)* and cancelling these terms, we get:

$$PFmean(lt) \times \frac{Constant}{Ref\ fdev} = cor(fdev, Pgrossdev) \times std(Pgrossdev) \times Constant \quad (29)$$

We can see here a MW quantity $std(PGrossdev)$ modified by the correlation between *fdev* and *Pgrossdev*, $cor(fdev, Pgrossdev)$ as well as by an undetermined but dimensionless term *Constant*. We can plug this into the settlement formula (22) to get a formula like the AEMC proposed settlement formula:

$$\begin{aligned} FPP(u, di) = CFLT(u, di) \times Ref\_Price(di) \times duration(di) \\ \times cor(fdev, Pgrossdev) \times std(Pgrossdev) \times Constant \end{aligned} \quad (30)$$

We need a robust basis for determining *Constant.* To do this, we can return to (15) and work through the algebra, assuming that all noise in the response has been eliminated. After aggregating (15) and assuming:

$$Pgrossdev(t) = Bias \times fdev(t) \quad (31)$$

i.e. that there is ideal performance, we note that the aggregated performance factor becomes a pure quadratic in *fdev*. As the frequency distribution is near normal with a standard deviation of *std(fdev)* essentially set by policy (currently it stands at about 25mHz), we can calculate the expected value of this quadratic over a long period. This turns out to be:

$$Bias \times std(fdev)^2 = std(Pgrossdev) \times std(fdev) \quad (32)$$

We can substitute this result into (23) to get:

$$PFmean(lt) \times \frac{Constant}{Ref\_fdev} = std(Pgrossdev) \times std(fdev) \times \frac{Constant}{Ref\_fdev} \quad (33)$$

Setting again *Ref_fdev* equal to *std(fdev)*, this simplifies to

$$PFmean(lt) \times \frac{Constant}{Ref} = std(Pgrossdev) \times Constant \quad (34)$$

Comparing this to (28) we see the result is the same except that the correlation term is missing or, rather, is equal to 1 under the assumption of ideal performance. This is an expected result. The correlation term discounts the performance measure to account for the imperfect relationship between *fdev* and *Pgrossdev.*

However, how do we interpret $std(Pgrossdev)$? The simplest way is by analogy to AC power flow, as set out in the table following. RMS is Root Mean Square, which is an equivalent concept to standard deviation for the zero-centred profiles which apply here.

Figure 31: Comparison of FDP and AC Power Flow concepts

#	FDP Variable	AC Power Flow Variable	Comment
1	*std(fdev)*	RMS Voltage	RMS and std are equivalent
2	*std(Pgrossdev)*	RMS current	RMS and std are equivalent
3	*cor(fdev,Pgrossdev)*	Power factor	Adjustment factor for mismatch
4	*std(fdev) x std(Pgrossdev)*	Apparent power	No adjustment for effective performance
5	*cor(fdev,Pgrossdev) x std(fdev) x std(Pgrossdev)*	Real power	Measure of effective performance
6	*peak(Pgrossdev)*	Peak current	Number of std relative to (2)
7	Gaussian (normally distributed) mean-reverting profiles	Sinusoidal profiles	Differing statistical behaviour

This analogy is instructive. In power engineering, the real power term (#5) in the table is a measure of the useful work done, as in the FDP case. However, the apparent power (#4) defines the current (analogous to the gross amount *std(Pgrossdev)* which must be accommodated in equipment that delivers useful work. To allow for this in the FDP case we must define:

$$Constant = \frac{1}{cor(fdev, Pgrossdev)} \quad (35)$$

Which from (30) gives a settlement formula. By setting:

$$Ref\_Price \; = \; Regulation\_Price \quad (36)$$

to use an appropriate price, we get:

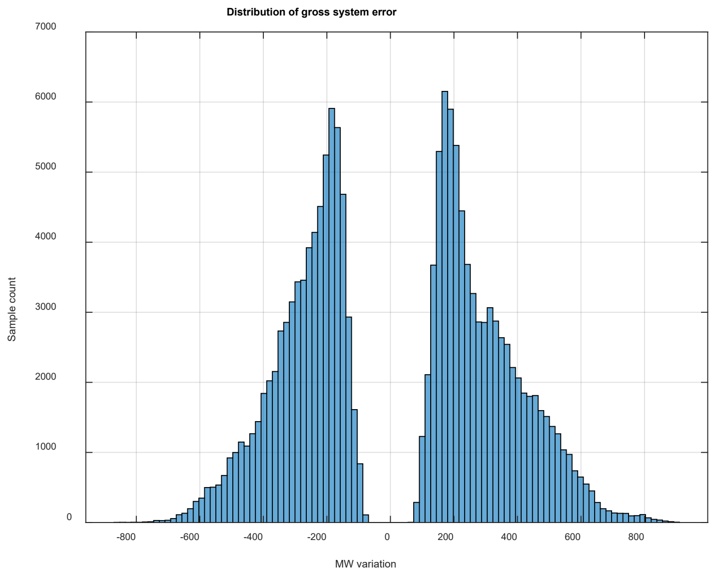
$$FPP(u, di) = CFLT(u, di) \times RegulationPrice \times duration(di) \times std(Pgrossdev) \quad (37)$$

We can look at this issue in another way. Figure 32 shows the distribution of gross system error for two weeks in September 2021. It is likely to be typical of any extended period.

If we use the formula (30) with *Constant = 1*, we are only covering the average (actually, root mean square) gross deviation. Th standard deviation or RMS of gross deviation is measured for the September sample period as 315MW, which is a somewhat beyond the peak away from zero on both sides. Examining the chart below, we see that the peak MW is about 800MW, representing a multiplier (number of standard deviations from zero) of about 2.5. This represents the “grossing up” required to set aside the required PFR and regulation capability. If we apply this multiplier as the value of *Constant,* we note that the two terms in (30) *cor(fdev,Pgrossdev)* and *Constant* we have just defined cancel out, as we measure *cor(fdev,Pgrossdev)* as about 0.4, giving a product of 2.5 x 0.4 = 1. In summary, we then return to the simple settlement formula (37).

While *std(Pgrossdev)* has been presented as a-long run value in this analysis, in practice it could vary over time. It should ideally be estimated in advance as it is intended to be a gross error that *could* be used but may not be. The actual outcome is measured and expressed in the ratio *CFLT*.

Figure 32: Distribution of Gross Deviations for two weeks in September 2021



C.3 Conclusions

The proposed FDP settlement formula in the general form proposed by the AEMC is as follows:

$$FPP(u,di) = CFLT(u,di) \times Regulation\_Price \times duration(di) \times std(Pgrossdev) \quad (38)$$

This is easily converted into a form strictly consistent with the AEMC proposed form by multiplying and dividing by *PFmean(di)* and re-arranging:

Here the term CLFT differs from the AEMC-defined CF because the denominator is a long run aggregate *PFmean(lt)* rather than one determined within the DI. Further the last MW terms is determined in advance as it is intended as an expected value, not an actual value, although this expectation could vary over time. Volatility in out-turn is expressed in CFLT.

This is easily converted into a form strictly consistent with the AEMC proposed form by multiplying and dividing by *PFmean(di)* and re-arranging:

$$FPP(u,di) = CF(u,di) \times Regulation\_Price(di) \times duration(di) \times std(Pgrossdev) \times \frac{PFmean(di)}{PFmean(lt)} \quad (39)$$

We can see from the last term how *std(Pgrossdev)* moves up and down with the call on performance. If there is no value in performing in a DI there is no payment. This contrasts with other measures for defining the last term such as *max(Pgrossdev, di*), which from Figure 32 above delivers an effective payment floor even when a raise or lower service is little used or not used at all.

For practical settlement, this formula can be much more simply expressed by breaking up our CF to get:

$$FPP(u,di) = PFmean(u,di) \times Regulation\_Price(di) \times duration(di) \times \frac{std(Pgrossdev)}{PFmean(lt)} \quad (40)$$

Or, alternatively

$$FPP(u,di) = PFMean(u,di) \times Regulation\_Price \times \frac{duration(di)}{std(fdev) \times cor(fdev, Pgrossdev)} \quad (41)$$

Note that the last term is a constant within the DI and could and should be constant over a whole billing period or more, only revised from time to time as needed.

For those who would wish to track and manage their performance within the DI, the formulae (41) and (42) above are consistent with the FDP formula (12),namely:

$$\frac{FDP}{Ref\_Price} = \frac{fdev}{Ref\_fdev} \times Constant \quad (42)$$

$$FDP(t) = fdev(t) \times Regulation\_Price(di) \times \frac{std(Pgrossdev)}{PFmean(lt)} \quad (43)$$

Or, alternatively, one directly derived from (12):

$$FDP(t) = fdev(t) \times Regulation\_Price(di) \times \frac{1}{std(fdev) \times cor(fdev, Pgrossdev)} \quad (44)$$

Note that everything but *fdev(t)* in both formulae is constant within the DI and known in advance. Note also that the FDP formula applies to all units in the system if the basic parameters are not changed. These should move only slowly or after a FOS adjustment and not normally within a billing period.

The denominator of the last term is reasonably stable; *std(fdev)* is about 25mHz at present although this target might change under FOS review, while *cor(fdev,Pgrossdev)* is currently about 0.4 but should gradually move closer to 1 over time.

So we calculate the current constant term in the FPP formula (41) as

Constant0 = 1/(12 x 0.025 x 0.4) = 100/12

In the FDP formula (43) there is no term duration(di), so for that formula:

Constant1 = 1/(0.025 x 0.4) = 100

With these parameters plugged in the settlement formula reduces to:

$$FPP(u,di) = PFmean(u,di) \times Regulation\_Price(di) \times 100/12 \qquad (45)$$

And the FDP formula becomes

$$FDP(t) = fdev(t) \times Regulation\_Price(di) \times 100 \qquad (46)$$

Example

For some unit *u* and dispatch interval *di*, suppose:

PFmean(u,di) =0.07 MW.Hz

Regulation_Price(di) = $60/MWh

For the dispatch interval *di* and unit *u*, the performance payment is then:

FPP(u,di) = 0.07 x 60 x 100/12 = $35

For some time *t* within the DI, suppose:

fdev(t) = 40 mHz

Then:

FDP(t) = 0.04 x 60 x 100 = $240/MWh

That price applies to the increment of deviation energy covered by that measurement interval.

APPENDIX D. PFR Correlation Analysis

D.1 Introduction

For Scenario 1 one of the performance metrics (PM) being considered for PFR is a Yes/No PM based on a threshold value for the correlation between deviation in system frequency (sensed by the DUID) and the MW deviation of the DUID from its dispatch target. The threshold is meant to determine if a DUID's performance is helpful, harmful or neither to the system. For metered units, helpful performance is rewarded and unhelpful performance results in a cost to the DUID.

The change in active power of a generator (MW) can be expressed as

$$\Delta P = P_{MAX} \times \frac{\Delta F}{50} \times \frac{100}{Droop\ (\%)} \tag{47}$$

Where:
ΔP is the active power change expressed in MW,
P_{MAX} is the Maximum Operating Level expressed in MW,
ΔF is the frequency deviation (in the document, beyond the deadband) expressed in Hz, and
Droop is expressed as a percentage.
Note that 50 Hz can be any other reference frequency.

The above terms are defined as in AEMO's Primary frequency response document to maintain familiarity.

D.2 A reasonable method for determining a correlation threshold

How a DUID responds to the change in system frequency is subject to a variety of factors. The actual response measured at any time period can be thought of as a random sample. Consistent with this view a correlation coefficient between the PM and MW deviation from target is a random variable subject to sampling error. To arrive at a non-arbitrary threshold value (magnitude) of the correlation coefficient, a hypothesis test can be conducted to determine if the measured correlation coefficient is significantly different than zero. The outcome of the test depends on the sample size (here the measurement or sampling period) and level of significance. Sample size 'n' depends on the length of the sampling period. For a 5-minute sample period *'n'* is 75 – as there are 75 4-second pairs of measurements. Similarly for a 30-minute period *'n'* is 450. If a significance level of 5% is selected, there is a 5% probability of concluding that the true correlation coefficient is not equal to zero (rejecting the null hypothesis as formulated below) when in fact it is zero. This is a standard interpretation of the Type I error.

Formally, the hypothesis test is:

Null hypothesis: $H_0\colon\ \rho = 0$

Alternate hypothesis: H_a: $\rho \neq 0$

This is a two-tailed test where ρ is the true value of the correlation coefficient.

D.3 Pearson's correlation

For two random variable x and y Pearson's correlation coefficient 'r' is given by the

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{48}$$

The random variable *'r'* follows a t-distribution with degrees of freedom equal to 'n-2'. This allows calculating a p-value associated with a value of the sample correlation 'r'.[1] If the p-value is below the level of significance 'α' (5% for example) the null hypothesis H_0: $\rho = 0$ is rejected and the DUID's performance can be categorised as helpful (if the magnitude is positive) and unhelpful (if the magnitude is negative). P-values higher than the selected level fail to reject the null hypothesis and lead to the conclusion that the true value of the correlation coefficient is not statistically different than zero. To give a sense of what this means for 'r' at the 5% level of significance, for '*n*'=75 the critical values correspond to values of 'r' of approximately ± 0.23 and for '*n'*=450 they are approximately ± 0.092. For the 30-minute sample period a value of 'r' that lies within the range ± 0.092 is not statistically different from zero (at the 5% level of significance). A value outside this range results in rejecting the null hypothesis. At the 1% level of significance the critical values are, somewhat higher, ± 0.296 and ± 0.121 for '*n'*=75 and '*n'*=450 respectively.

Determining a threshold based on 5-minute sample periods, while attractive at first sight, results in the decision to reject or not reject the null hypothesis in 5-minute periods changing too frequently to be sensible. Longer periods are more attractive both from system performance and statistical viewpoint. However there are other considerations. A period such as a billing week may be judged too long from the perspective of a 5-minute settlement period. This indicates a limitation of this approach. The result would be more robust the longer the period that is used but subjective judgment may favour shorter periods such as a 5-minute or a 30-minute period.

The correlation coefficient was calculated on a 5-minute and 30-minute basis over the one-week period from 1 to 7 September 2021 for two generators. A chart for each of the 5-minute and 30-minute values for ER01 and KPP1 is shown below. Both charts show that the sample correlations cross the critical values very frequently. For example, at the 5% level of significance the 5-minute correlation coefficients cross the critical values 587 and 542 times for ER01 and KPP1 respectively. At 30-minutes they cross 194 and 212 times. The number of crosses reduces going from 5- to 30- minutes but by a factor of less than 6. This indicates that even at 30-minutes the "frequency" with which the sample correlation values cross the critical values is still quite high.

Figure 33: 5-minute correlation coefficients for ER01 & KPP1 over a 1-week sample period

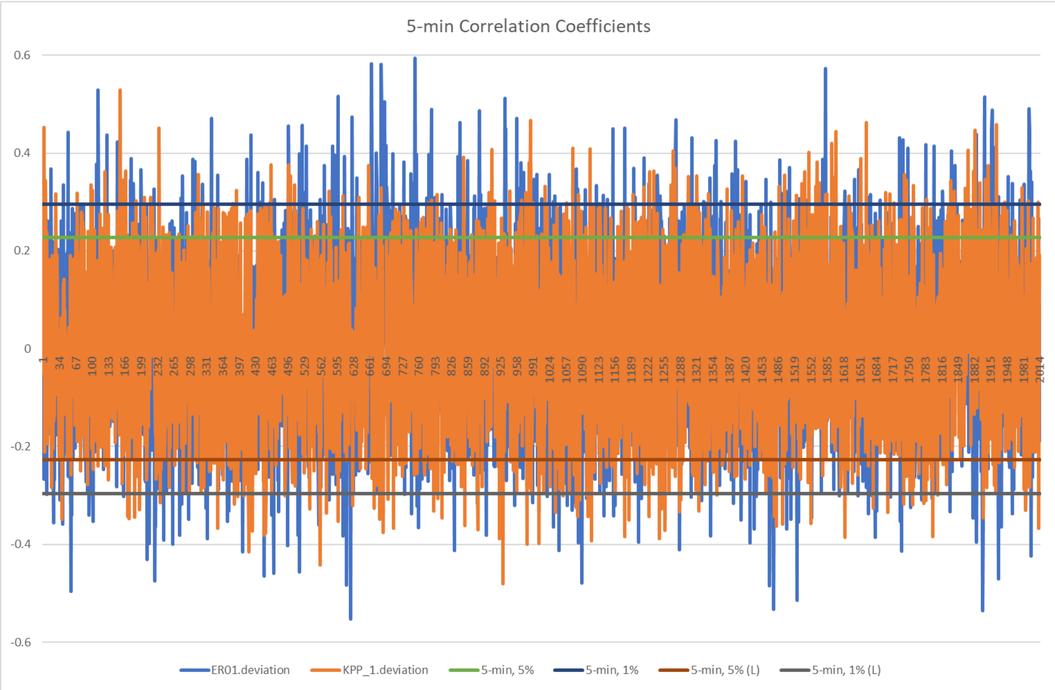
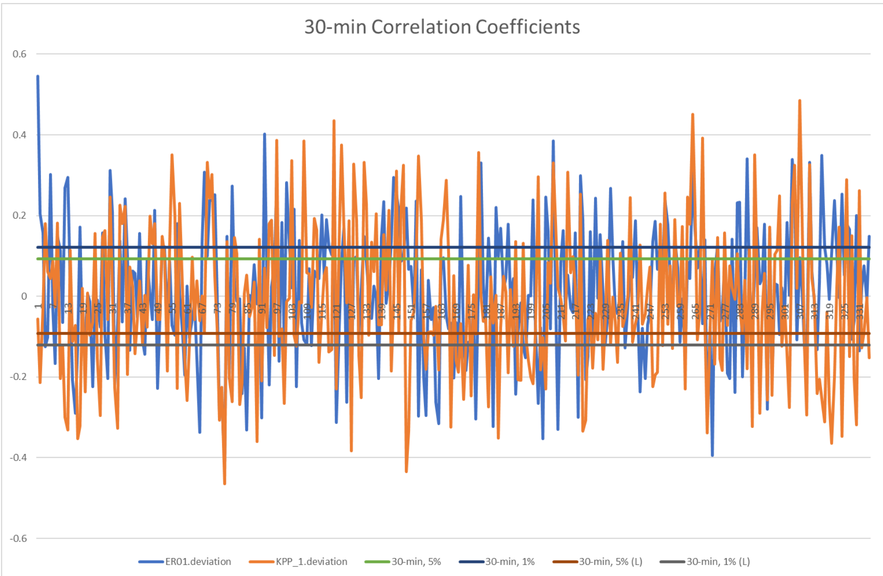
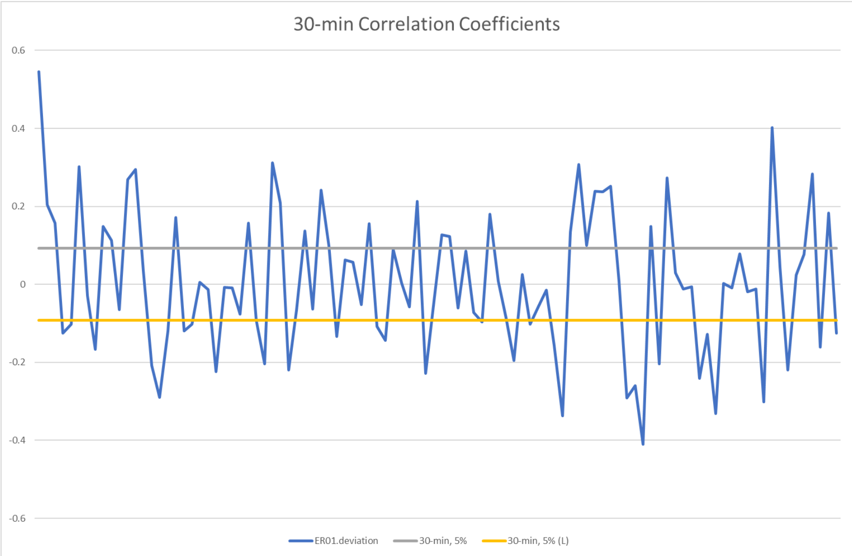


Figure 34: 30-minute correlation coefficients for ER01 & KPP1 over 1-week sample period



To better visualise the results Figure 3 zooms in on the first 100 30-minute correlation sample values in the sampling period for ER01. We see the sample correlation values crossing from the positive (helpful) to the negative (unhelpful) zones quite frequently.

Figure 3: 5-minute correlation coefficients for ER01 & KPP1 first 100 sample values in the sampling period 5% level



The mean of the frequency metric over the one-week period from 1 to 7 September 2021 is -0.0003 Hz and its standard error is 0.00007 which indicates that the mean is not statistically different from zero. This aligns with expectations.

D.4 Regression analysis

Equation (47) can be rearranged as follows:

$$\Delta P = \Delta F \times \frac{100}{50} \times \frac{P_{MAX}}{Droop\ (\%)} \tag{49}$$

If we consider *'P_{MAX}'* and *'Droop (%)'* to be constant this is equivalent to:

$$\Delta P = Constant \times \Delta F \tag{50}$$

If we regress the unit's real power deviation (in MW) over the frequency deviation performance metric (the negative of the system's frequency deviation) we are essentially regressing a variant of the above equation (50), which is given as equation (51) below

$$\Delta P = \beta_0 + \beta_1 \Delta F + \varepsilon \tag{51}$$

Data for the one-week period from 1 to 7 September 2021 from a small sample of generators were regressed but the resulting coefficient of determination (r^2) was found to be quite low, refer Table 6-10.

Table 6-10: Coefficient of correlation and coefficient of determination of sample plant

	ER01	KPP1	GSTONE2	GSTONE3	GSTONE5
Correlation coefficient	0.451	0.250	0.095	0.127	0.158
Coefficient of determination	0.203	0.062	0.009	0.016	0.025

The scatter plots below show a high scatter and visually confirm the low coefficient of determination of the regression models.

Figure 35: Scatter plot of MW deviation vs Frequency metric over 1 week - ER01

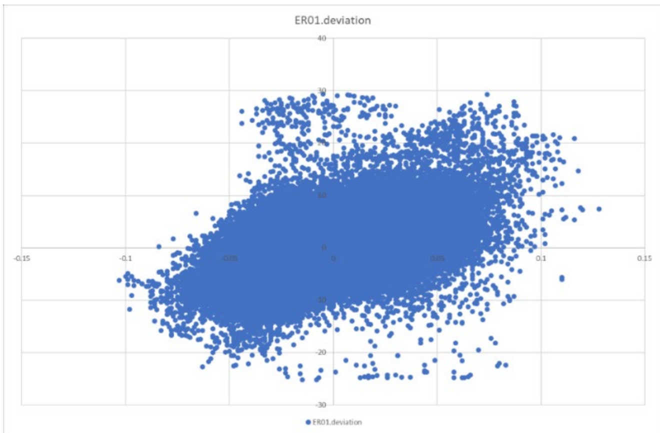
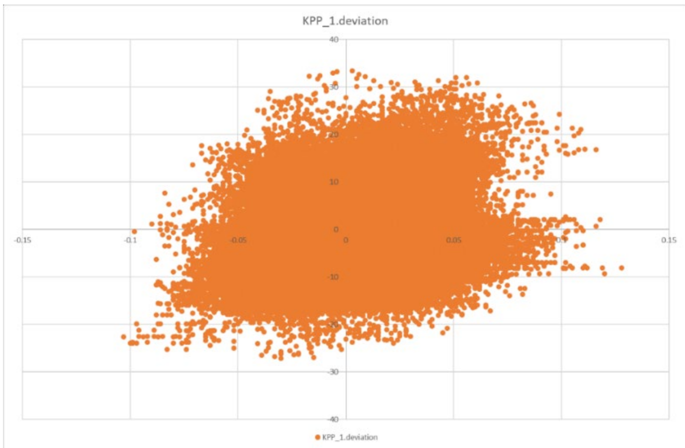


Figure 36: Scatter plot of MW deviation vs Frequency metric over 1 week - KPP1



The coefficient of regression β1 in equation (51) can be expressed as in equation (52)

$$Coefficient\ of\ Regressions\ \beta_1 = \frac{100}{50} \times \frac{P_{MAX}}{Droop\ (\%)} \tag{52}$$

Equation (52) can be used to calculate the droop of a plant whose MW deviation is modelled. The resulting values of droop are larger than expected.[3] This is due to the large variance of the error term referred to above.

D.5 Improving the regression model

To improve the goodness of fit categorical variables (also referred to as dummy variables) were introduced to differentiate between the region below the lower limit of the deadband, the region above the upper limit of the deadband and the region within the deadband (± 0.015). The improvement was quite small. The variance of the error term is large, and the simple regression model (even after the addition of the categorical variables) does not capture well the behaviour of the data. This motivated a different strategy to improve model fit.

In equation (47) '*ΔF*' is defined as the frequency deviation (in the document, beyond the deadband) expressed in Hz. This suggests that a unit that has its dead band set at the maximum allowed deadband limits would not have a primary frequency response in accordance with its droop settings while the frequency deviation remains within the limits of the frequency deadband.[4] Furthermore, the response can be different in the region on both sides of the frequency deadband. To explore this the data was divided into three subsets: within the frequency deadband, above and below the deadband. Regression of the last two subsets was conducted separately. Visual inspection of the scatter plot of ER01 in Figure 35 also lends support to exploring this line of analysis.

Formally we re-estimate regression equation (51) except this time we use only the data subset that corresponds to '*ΔF*' > 0.015. We also subtract 0.015 from the frequency metric deviations[5] and force the vertical intercept to pass through zero. This has the effect of aligning the '*ΔF*' that is expected to drive a MW deviation with the definition used in relation to the droop settings. We repeat the same process for the data that corresponds to '*ΔF*' < - 0.015. The slight difference is that we now add 0.015[6] to and force the intercept to be zero. A summary of the results is provided in Table 6-11. The improvement in goodness of fit is mixed. For ER01 the improvement is mostly in the >0.015 subset. KPP1 improved for the subset <-0.015 but deteriorated (much lower coefficient of determination) for the subset > 0.015. GSTONE units 2,3 &5 improved for the subset > 0.015 but deteriorated for the subset <-0.015 and surprisingly this subset resulted in a change of sign of the regression coefficient, suggesting that these units were unhelpful to system frequency in this subset (<-0.015).

Table 6-11: Coefficient of correlation and coefficient of determination of sample plant – includes regression on subsets

	ER01	KPP1	GSTONE2	GSTONE3	GSTONE5
Subset above the upper limit of the deadband (>0.015)					
Regression coefficient	147.012	27.075	58.769	63.857	72.865
Coefficient of determination	0.309	0.005	0.084	0.116	0.136
Subset below the lower limit of the deadband (<-0.015)					
Regression coefficient	128.004	205.725	-14.696	-11.715	-10.200
Coefficient of determination	0.220	0.254	0.004	0.004	0.003
Within the deadband					
Regression coefficient	80.058	62.927	13.852	15.770	19.872
Coefficient of determination	0.203	0.062	0.009	0.016	0.025

Dividing the data into subsets has resulted in some improvement and indicated possible differences in the response of plant when frequency is above the upper limit of the deadband compared to when it is below the lower limit. But caution needs to be exercised when basing conclusions on only a small subset of coal fired plant.

The low goodness of fit indicates large variation in the error term. To remedy this, more factors that reflect the behaviour of the generating units would need to be introduced into the model. However, these factors depend on the operating characteristics of individual plant and, in our view, a single model cannot reliably apply to all generation plant.

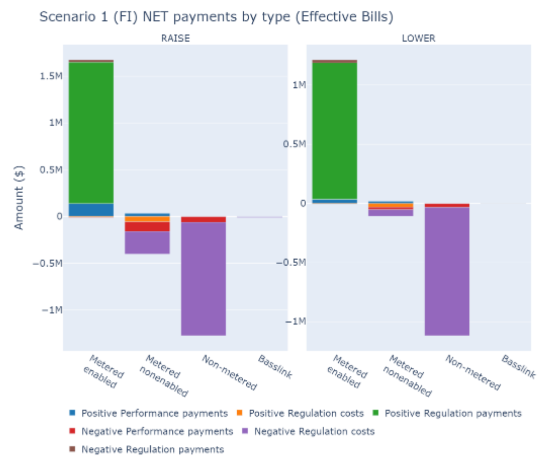
D.6 Conclusion

A hypothesis test can be used to determine a threshold value of the correlation coefficient. However, the objectivity of the test is diminished as the choice of the period to calculate the sample correlation coefficient is based on judgement (subjective). As the regressions show, the regression relationship is quite weak. This indicates that the estimate of the coefficient of regression of the relationship between the change in active power (MW deviation) and frequency deviation is impacted by high variability in the error term as the simple regression model does not control for other factors that drive the plant's behaviour. Other factors depend on plant type and individual plant operating characteristics. In our view a single model cannot reliably describe all generation plant.

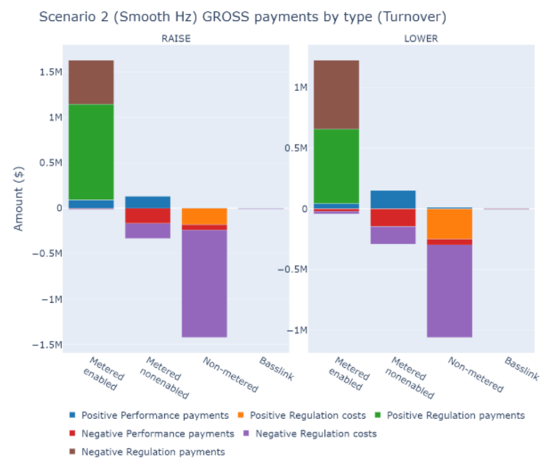
APPENDIX E. Scenario results over two weeks

Figure 37 : Base scenario full settlement results over two weeks

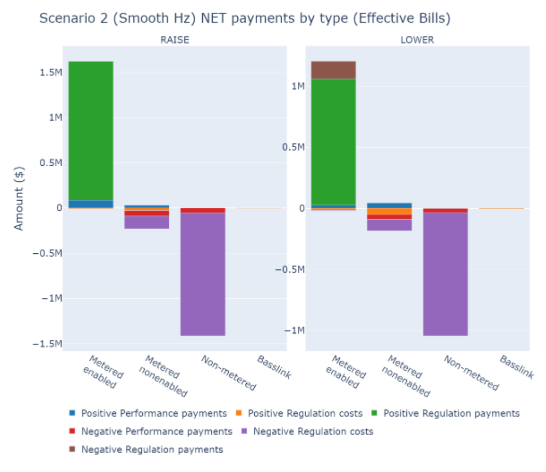
Scenario 1 – Frequency Indicator (FI): Net



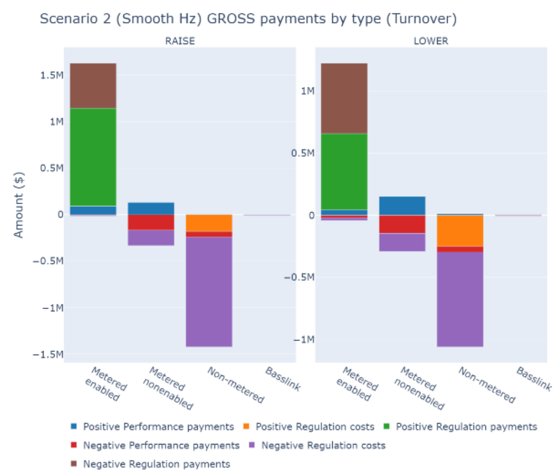
Scenario 1 – Frequency Indicator (FI) - Gross



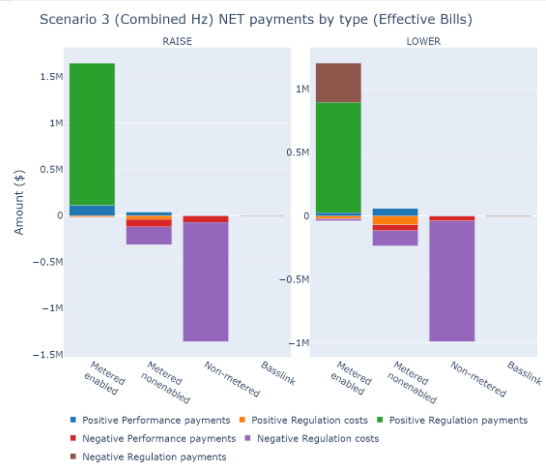
Scenario 2 – Smoothed Hz - Net



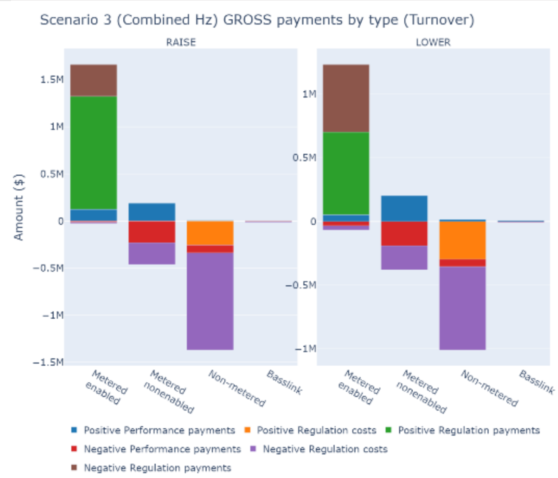
Scenario 2 – Smoothed Hz - Gross



Scenario 3 – Combined Hz - Net



Scenario 3 – Combined Hz - Gross



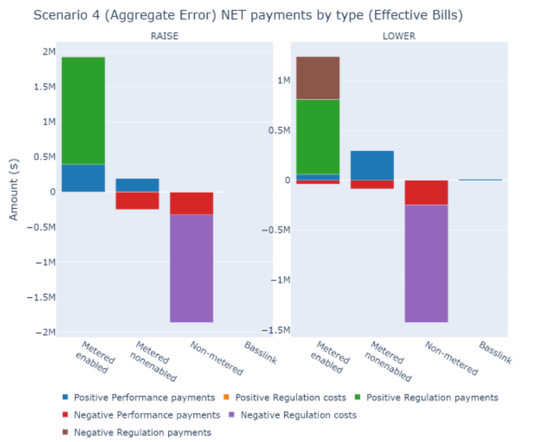
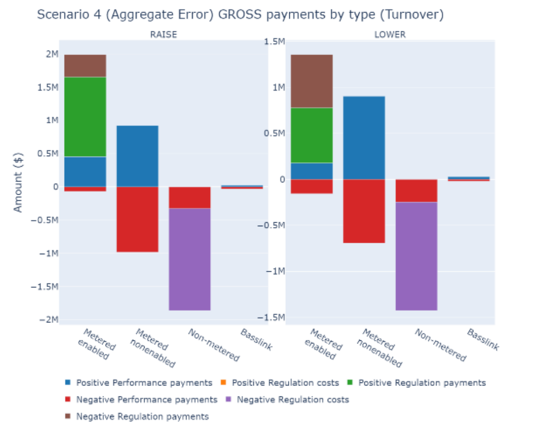
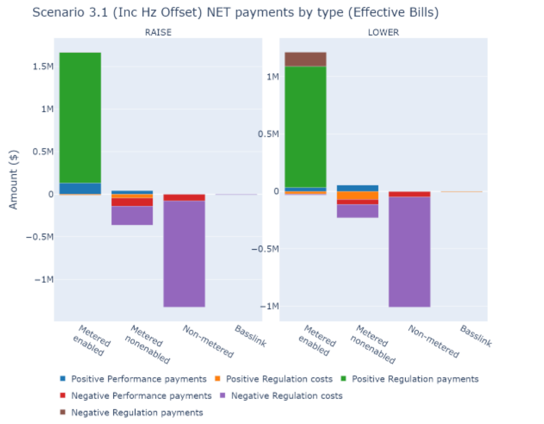
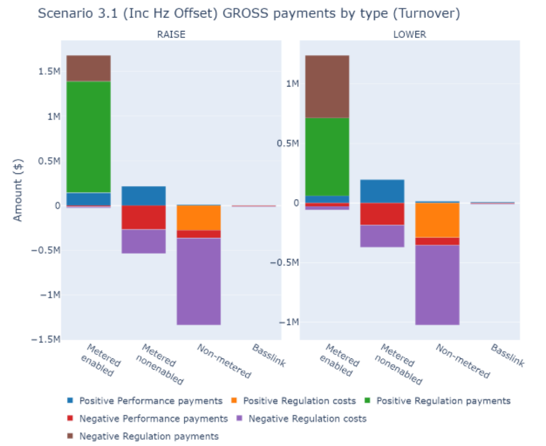
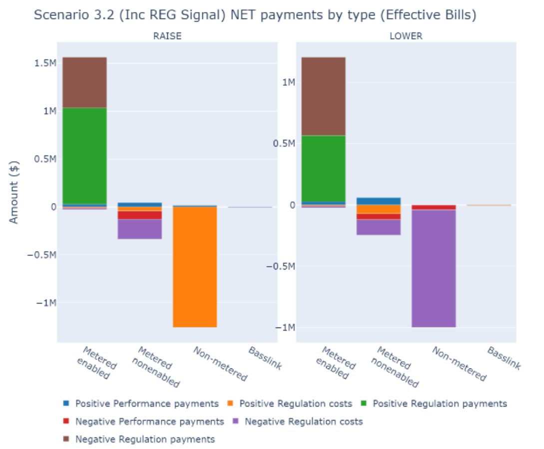
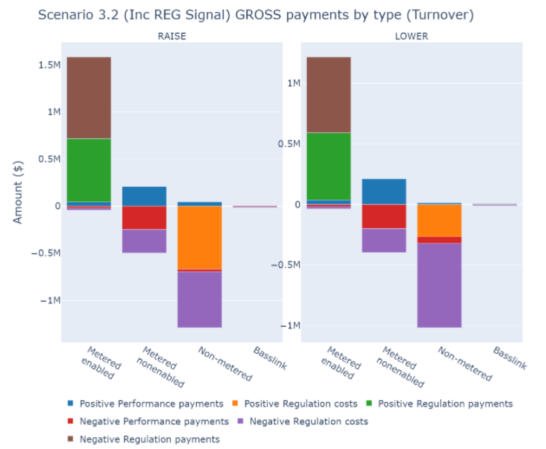
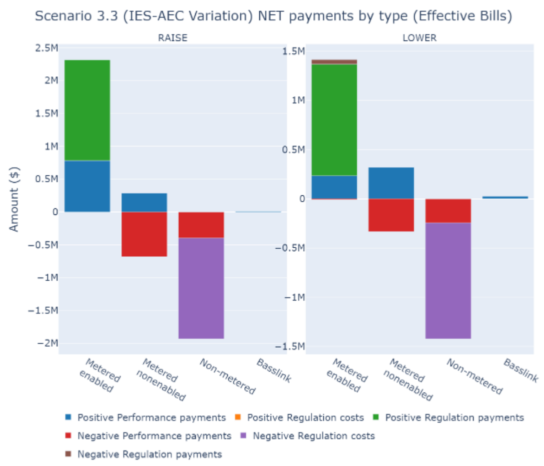
Scenario 4 – Gross Dispatch Error - Net	Scenario 4 – Gross Dispatch Error - Gross
	

Figure 38: Additional Scenario results based on Scenario 3 over two weeks

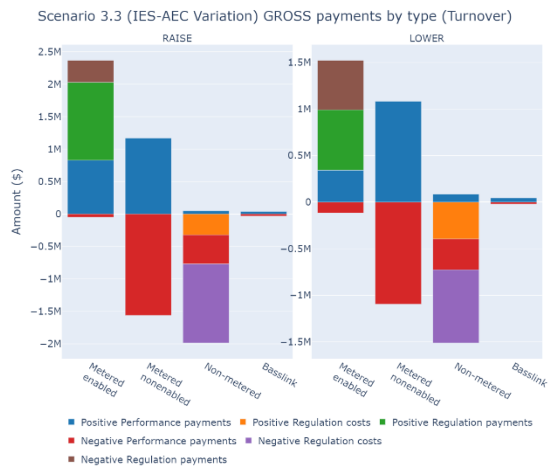
Scenario 3.1: Net	Scenario 3.1: Gross
	

Scenario 3.2 – Net	Scenario 3.2 – Gross
	

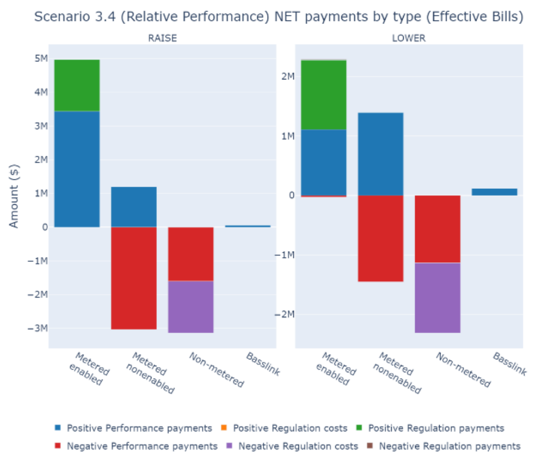
Scenario 3.3: Net



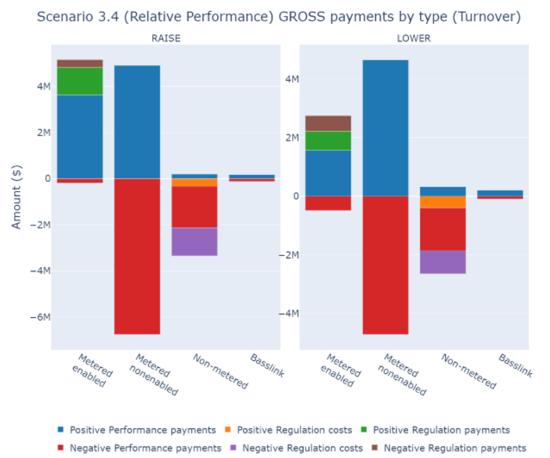
Scenario 3.3: Gross



Scenario 3.4: Net



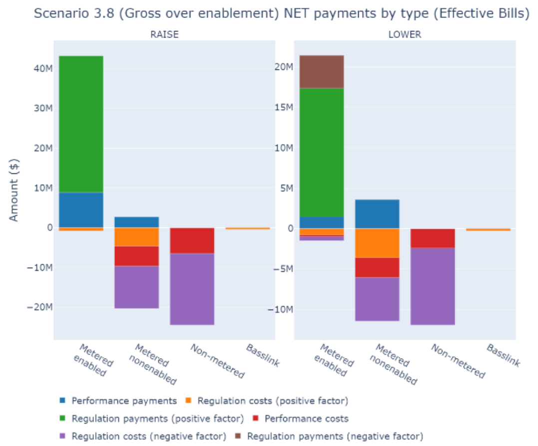
Scenario3.4: Gross



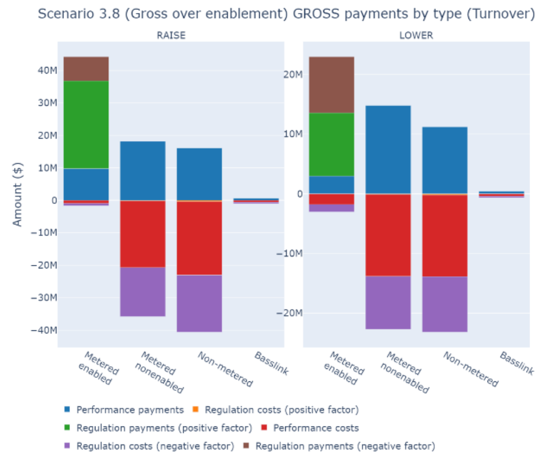
APPENDIX F. Scenario Results over Eleven Months

Figure 39: Long period summary results per scenario

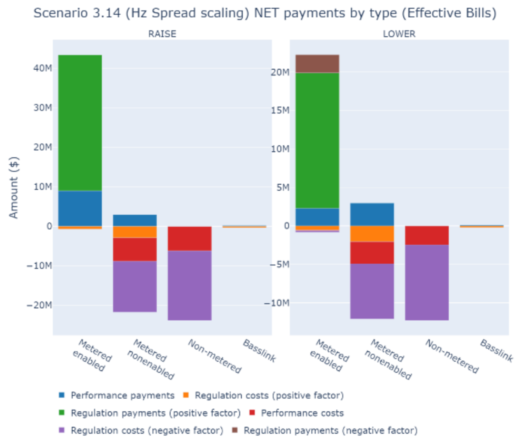
Scenario 3.8– Net: Gross system error scaling



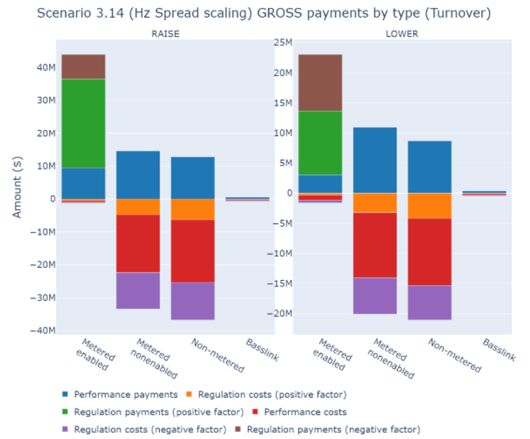
Scenario 3.8 –Gross: Gross system error scaling



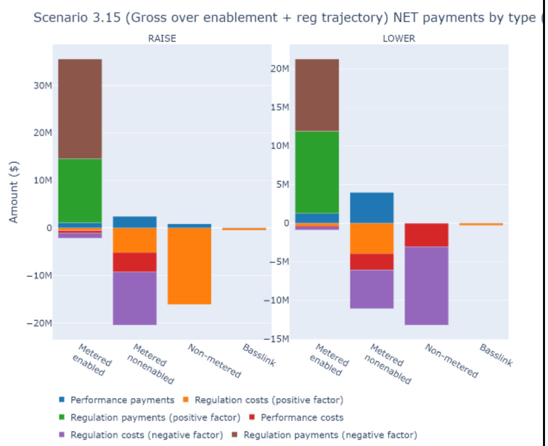
Scenario 3.14 – Net: Hz spread scaling



Scenario 3.14 – Gross: Hz spread scaling



Scenario 3.15 – Net: Including AGC reg trajectory in reference trajectory



Scenario 3.15 – Gross: Including AGC reg trajectory in reference trajectory

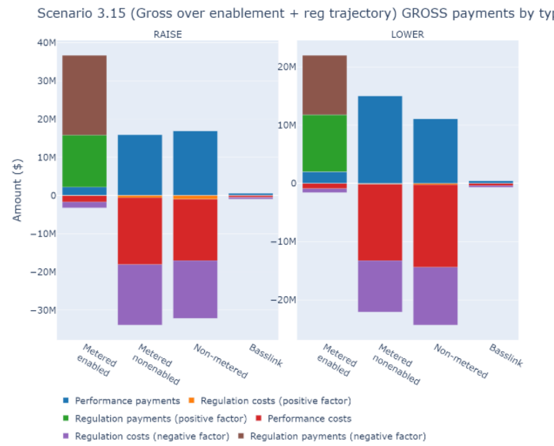
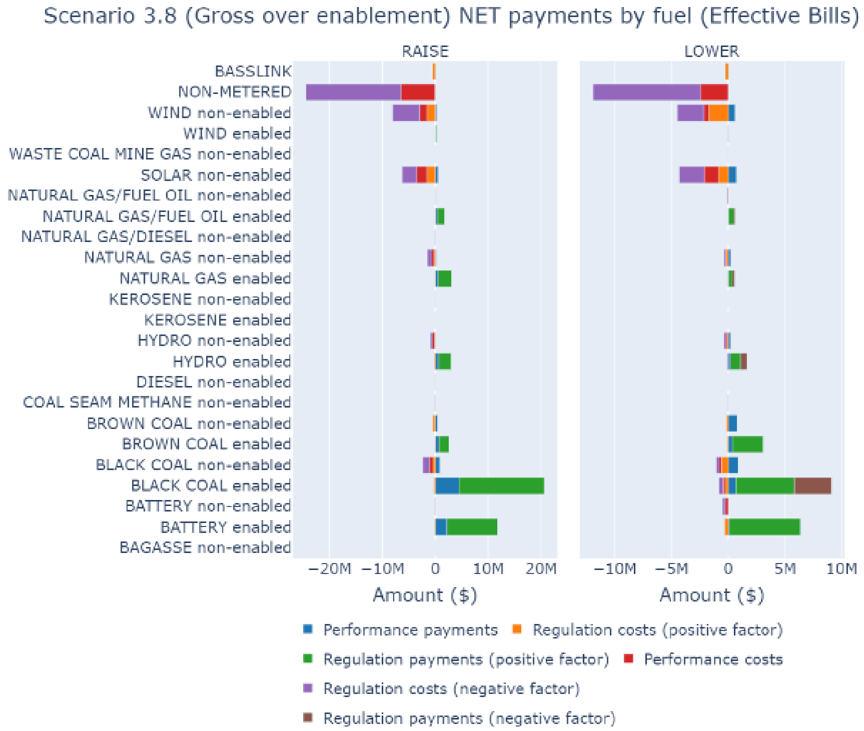
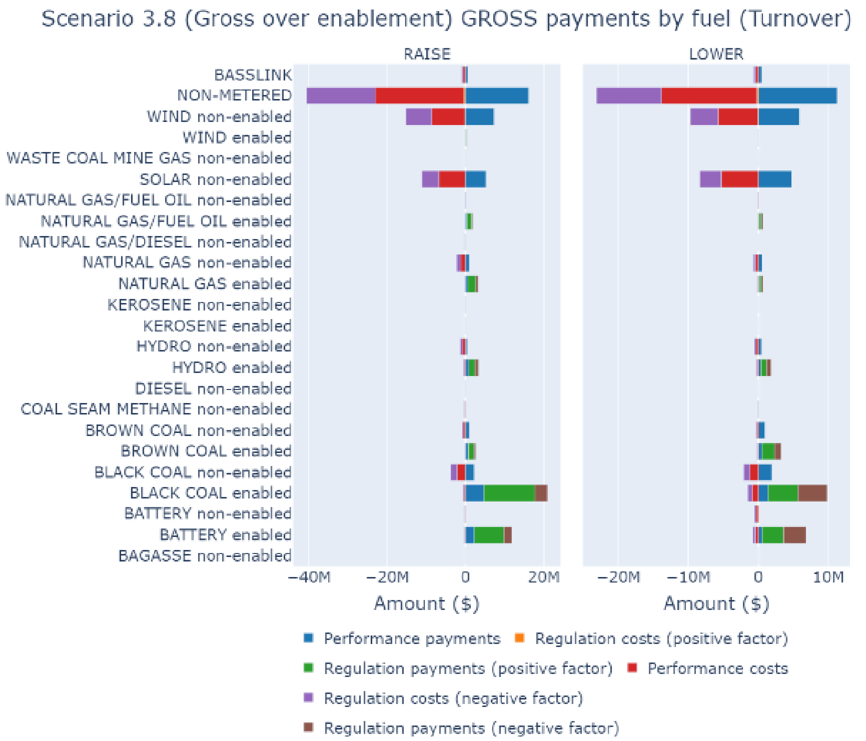


Figure 40: Summary charts by fuel type and enablement

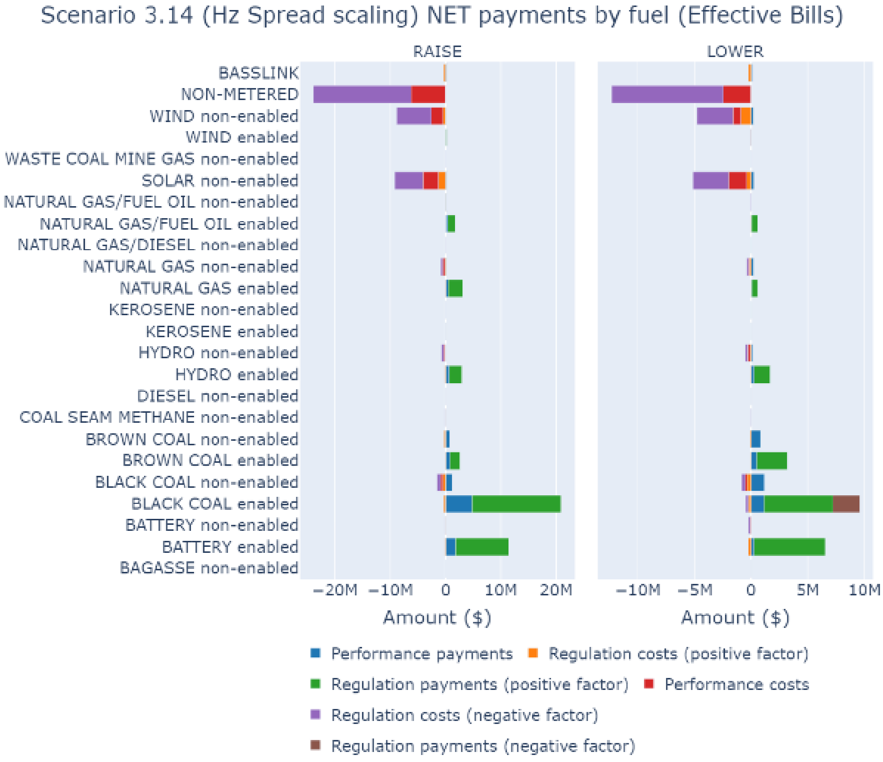
Scenario 3.8 – Net payments by fuel



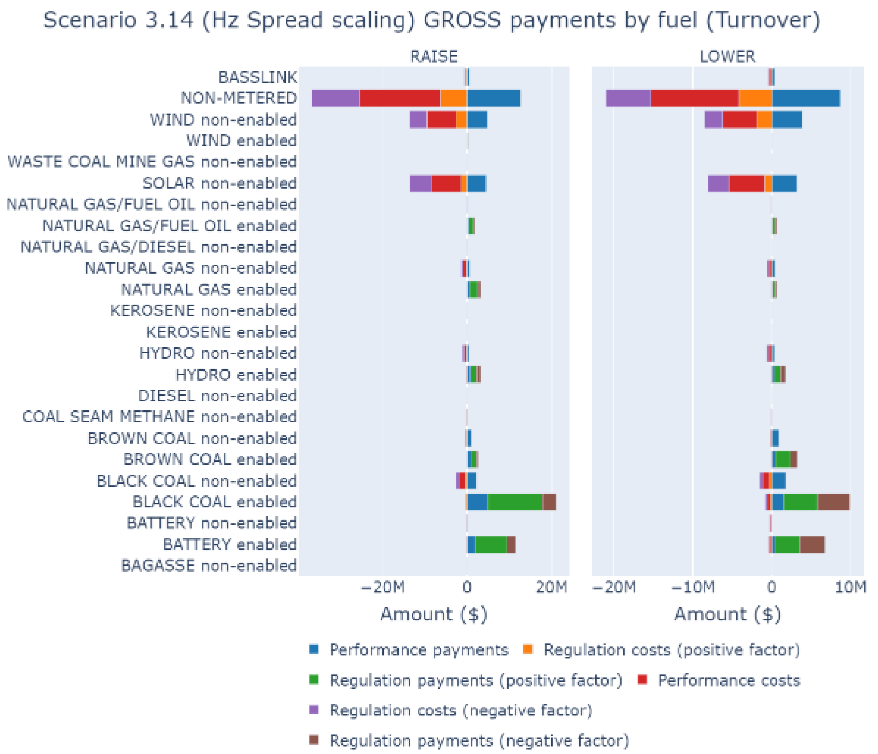
Scenario 3.8: Gross payments by fuel

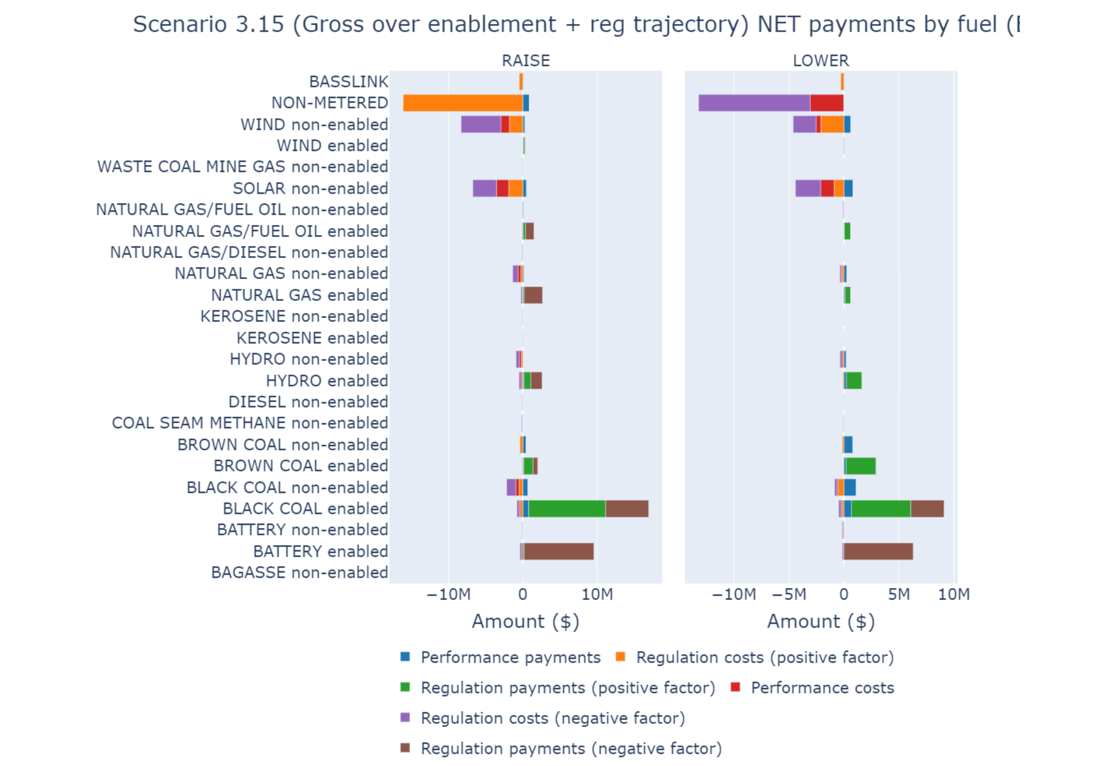


Scenario 3.14 – Net payments by fuel



Scenario 3.14: Gross payments by fuel



Scenario 3:15: Net payments by fuel

Scenario 3.15: Gross payments by fuel