Performance Accuracy Scores in CAISO and MISO Regulation Markets: A Comparison based on Real Data and Mathematical Analysis

Ashkan Sadeghi-Mobarakeh, *Student Member, IEEE*, Hamed Mohsenian-Rad, *Senior Member, IEEE*

***Abstract*—This paper investigates the fundamental differences between how California Independent System Operator (ISO) and Mid-continent ISO calculate performance accuracy scores in their performance-based regulation markets. Both ISOs tend to follow the Federal Energy Regulatory Commission (FERC) order 755 to pay regulation resources - whether conventional generators or distributed and demand side resources - based on their actual performance. The advantages and disadvantages of each method is systematically explained. First, real-world ISO data is used to show that there may exist major differences between these two methods in scoring accuracy under similar regulation performance scenarios. Next, the root-causes for the observed differences are studied mathematically. Finally, some suggestions are made to improve these scoring methods; should these or other ISOs seek to refine their scoring formulas.**

***Keywords*: Performance-based regulation market, performance score, mileage, California ISO, Mid-continent ISO.**

I. Introduction

Federal Energy Regulatory Commission (FERC) Order 755 requires Independent System Operators to develop *pay-for-performance* protocols in regulation markets to compensate regulation resources based on their *actual performance*. Such a payment, a.k.a, *mileage payment*, must reflect the regulation resource's speed and accuracy in following the Automatic Generation Control (AGC) dispatch regulation signal [1]. The goal is to compensate fast-response resources, such as aggregated and autonomous demand side resources, including electric vehicles, or batteries, based on their actual values.

California ISO (CAISO) and Midcontinent ISO (MISO) have both adopted this new market mechanism. They both calculate the mileage payment as a product of three terms:

$$\left\{\begin{matrix}\text{Mileage}\\\text{Payment}\end{matrix}\right\} = \left\{\begin{matrix}\text{Actual}\\\text{Mileage}\end{matrix}\right\} \times \left\{\begin{matrix}\text{Mileage}\\\text{Price}\end{matrix}\right\} \times \left\{\begin{matrix}\text{Performance}\\\text{Score}\end{matrix}\right\}. \tag{1}$$

The *actual mileage* is the up/down actual movement of the resource to follow the AGC dispatch signal. The *mileage price* is determined by the market. The *performance score* is a coefficient that evaluates the performance of the resource in terms of following the AGC dispatch signal. The last item, i.e., the performance score, is the matter of our focus in this paper because it is a key factor in implementing FERC Order 755 to reward resources based on their performance [2].

CAISO and MISO use different formulations for their performance score. Each ISO has its own considerations for its choice. Our focus is *not* on questioning the metrics used by these two ISOs. Instead, we seek to understand *how* these two metrics result in different implications. Our study is motivated by some recent ISO reports that raise concerns about their own performance scoring methods. For example, as noted in [3, p. 12], CAISO is interested in refining its scoring method based on comparisons with other ISOs: "The CAISO reviewed different methods for accounting for accuracy used by other independent system operators and regional transmission operators in their Order 755 market designs. Alternative approaches adopted by the Midcontinent Independent System Operator (MISO), the New York Independent System Operator (NYISO) and PJM Interconnection may provide some guidance on how to refine the ISOs performance metric."

The contributions in this paper are as follows:

1) A systematic approach is taken to compare the performance scoring methods used by CAISO and MISO and to identify the advantages/disadvantages of each method.
2) By using real-world ISO data, it is shown that these two methods may result in significantly different scores for similar regulation performance scenarios.
3) Our study required doing a detailed mathematical analysis of the MISO performance accuracy score, which to the best of our knowledge, is done for the first time.
4) Some recommendations are made to refine and possibly improve these two performance scoring methods.

II. Two Performance Scoring Methods

The performance score in CAISO is a number between 0 and 1, which we denote by PS_{CAISO}. It is calculated once for each market interval that takes 15 minutes. The method of calculation is explained in [4]. Mathematically, we can write:

$$\text{PS}_{\text{CAISO}} = \left[1 - \frac{\sum_{\tau=1}^{T} \mid s[\tau] - y[\tau] \mid}{\sum_{\tau=1}^{T} s[\tau]}\right]^{+}, \tag{2}$$

where at each time slot τ of length four seconds, $s[\tau]$ and $y[\tau]$ denote the AGC setpoint and the mechanical output of the regulation resource. The fraction in (2) is a normalized measure of performance *inaccuracy* in following the AGC setpoints. Thus, 1 minus the fraction is used to obtain a performance *accuracy* measure. Note that, $[x]^{+} = \max\{0, x\}$; and $T = 15 \times 60/4 = 225$ denotes the number of time slots.

The performance score in MISO is also a number between 0 and 1. We denote it by PS_{MISO}. It is calculated once for each market interval that takes five minutes. The AGC setpoints are sent once every four seconds. MISO first calculates the *actual response*, which is the accumulation of changes in the output of the resource in response to the AGC setpoints, where a positive value indicates a move *towards* the AGC setpoint and a negative value indicates a move *away* from the AGC setpoint. The expected mileage is the desired movement towards the AGC setpoints starting at the mechanical output of the resource at the beginning of the market interval [5]. We can write:

$$\text{PS}_{\text{MISO}} = \left[\frac{\sum_{\tau=1}^{T} \mid s[\tau] - y[\tau-1] \mid - \mid s[\tau] - y[\tau] \mid}{\sum_{\tau=1}^{T} \mid s[\tau] - s[\tau-1] \mid}\right]^{+}, \tag{3}$$

The authors are with the ECE Department, UC Riverside, CA, USA. This work is supported by NSF Grants 1253516, 1307756, and 1711944. The corresponding author is H. Mohsenian-Rad. E-mail: hamed@ee.ucr.edu.

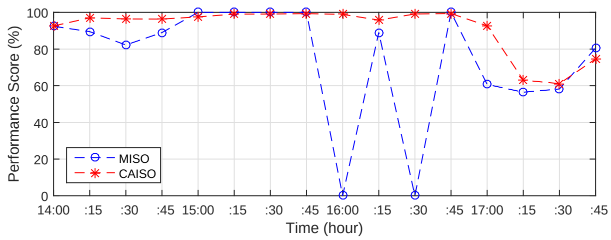

Fig. 1. Performance scores of the real-world ISO regulation data in [6].

TABLE I
PERFORMANCE SCORES FOR THE FOUR CASES IN FIG. 2.

Score	Case1		Case2		Case3		Case 4	
	Resources		Resources		Resources		Resources	
	1	2	1	2	1	2	1	2
PS_{CAISO}	0.72	0.72	0.67	0.89	0.76	0.60	0.76	0
PS_{MISO}	0.67	0	1	0	0	0.25	0.67	0.67

where $s[0] = y[0]$. Note that, the model in (3) does not consider the ramp constraints. Adding the ramp constraints does not change the main conclusions in this paper; it only unnecessarily complicates the notations and equations. To consider the ramp limits, one should change the term inside the summation in the denominator in (3) to $|s[\tau] - r[\tau-1]| - |s[\tau] - r[\tau]|$, where $r[\tau]$ is the ramp setpoint and $r[0] = y[0]$. Also note that, MISO does not distinguish the regulation mileage and energy mileage as far as the performance score is concerned [5]. However, since our focus here is on performance scoring methods of MISO and CAISO, we do not discuss this issue.

In (2) and (3), the *error*, i.e., the distance from AGC signal, and the *movements* towards AGC setpoints at each four seconds interval are the measures for evaluating a good resource, respectively. Next, we follow the same philosophy in evaluating the performance scores used by each of the ISOs.

III. Comparison of the Two Methods

Consider the real-world ISO data in [6], comprising AGC setpoints and mechanical outputs of a regulation resource. Fig. 1 compares the corresponding PS_{CAISO} and PS_{MISO}. Sixteen market intervals, each taking 15 minutes, are analyzed. Each time slot takes four seconds. We can see that the performance scores are very different. At certain intervals, such as 16:00-16:15 and 16:30-16:45, the two scores are contradictory, where one score is 0% and the other one is almost 100%.

Next, we study four additional representative test cases, as shown in Fig. 2. Each market interval takes 20 seconds. The performance scores are given in Table I. In Cases 1-3, two resources with outputs $y_1[\tau]$ and $y_2[\tau]$ tend to follow the same AGC signal $s[\tau]$. In Case 4, $y_1[\tau]$ follows $s_1[\tau]$ and $y_2[\tau]$ follows $s_2[\tau]$. The regulation resource output y could be from a conventional generation or a distributed resources, such as aggregated and autonomous demand response units [7]. In Case 1, both resources make equal absolute errors. Thus, PS_{CAISO} is the same for both resources. However, Resource 1 always moves in the *direction* of the AGC signal while Resource 2 moves in the *opposite direction* of the AGC signal. Hence, PS_{MISO} rewards Resource 1 and penalizes Resource 2. In Case 2, Resource 1 does *not* follow the AGC signal,

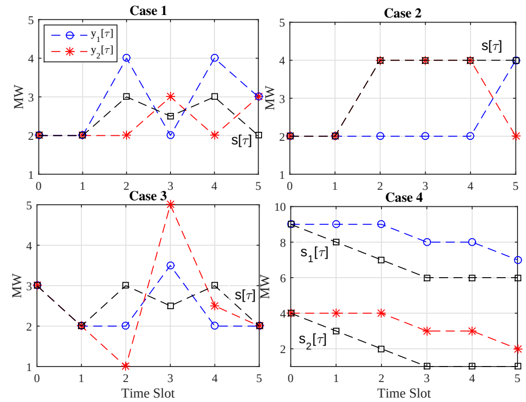

Fig. 2. Four case studies to study performance scores, where $T = 5$.

except until the last time slot. Resource 2 *does* follow the AGC signal, except in the last time slot. While PS_{CAISO} is higher for Resource 2, PS_{MISO} is drastically lower for Resource 2.

In Case 3, Resource 2 makes large errors when following the AGC setpoints and thus PS_{CAISO} favors Resource 1. However, PS_{MISO} favors Resource 2. Interestingly, similar patterns can be seen in Fig. 1. The total error is higher yet the MISO score is larger in interval 14:00-14:15 than interval 14:15-14:30.

In Case 4, we have $y_1[\tau] = y_2[\tau] + 5$ MW and $s_1[\tau] = s_2[\tau] + 5$ MW. Thus, the relativeness of $y_1[\tau]$ to $s_1[\tau]$ is the same as the relativeness of $y_2[\tau]$ to $s_2[\tau]$. Assuming that the ramp rates are the same, the two resources perform similarly in following the AGC signal. They also create equal total regulation errors, at 8 MW. However, PS_{CAISO} is very different for the two resources. PS_{CAISO} is *very sensitive* to the magnitude of the AGC setpoints. PS_{MISO} is reasonably similar for both resources. This is because the magnitude of the AGC setpoints forms a *bias* in the denominator in (2), even if the mileage values are exactly the same as for both resources.

IV. Underlying Causes in Cases 2 and 3

The following Theorem is used to explain Cases 2 and 3.

Theorem 1. *The numerator in (3), i.e., the so-called actual movement, can be written in the following equivalent form:*

$$\sum_{\tau=1}^{T} \Big| s[\tau] - s[\tau-1] \Big| - 2 \sum_{\tau \in \Psi} \Big| s[\tau] - s[\tau-1] \Big| - 2 \sum_{\tau \in \Phi} \Big| s[\tau-1] - y[\tau-1] \Big| - \Big| s[T] - y[T] \Big|, \quad (4)$$

where Φ is the set of all time slots $\tau = 1, \ldots, T$ such that either $s[\tau] \geq y[\tau-1] > s[\tau-1]$ or $s[\tau] \leq y[\tau-1] < s[\tau-1]$. Ψ is the set of all time slots such that either $y[\tau-1] > s[\tau] > s[\tau-1]$ or $y[\tau-1] < s[\tau] < s[\tau-1]$. The set of all those time slots that do not *belong to Φ and Ψ is defined by Ω.*

Proof: After adding and subtracting $s[\tau-1]$, we can rewrite the *first* term in the numerator in (3) as follows:

$$\sum_{\tau=1}^{T} \Big| s[\tau] - s[\tau-1] + s[\tau-1] - y[\tau-1] \Big|$$
$$= \sum_{\tau \in \Phi} \Big| s[\tau] - s[\tau-1] \Big| - \Big| s[\tau-1] - y[\tau-1] \Big| +$$

$$\sum_{\tau\in\Psi}\Big|\, s[\tau-1]-y[\tau-1]\,\Big| - \Big|\, s[\tau]-s[\tau-1]\,\Big| + \sum_{\tau\in\Omega}\Big|\, s[\tau-1]-y[\tau-1]\,\Big| + \Big|\, s[\tau]-s[\tau-1]\,\Big|, \tag{5}$$

We can rewrite the *second* term in (3) as

$$\sum_{\tau=0}^{T-1}\Big|\, s[\tau]-y[\tau]\,\Big| - \Big|\, s[0]-y[0]\,\Big| + \Big|\, s[T]-y[T]\,\Big| = \sum_{\tau=1}^{T}\Big|\, s[\tau-1]-y[\tau-1]\,\Big| + \Big|\, s[T]-y[T]\,\Big|, \tag{6}$$

where we used the fact that $s[0]=y[0]$. By subtracting (6) from (5), we can rewrite the numerator in (3) as (4). ■

One can use the same methodology as in the above proof and go through similar but unnecessarily more complex equations, so as to expand Theorem 1 to the case with the presence of the ramp constraints. The conclusions will remain the same. Next, we explain the implications of the results in Theorem 1.

On one hand, the first term in (4) in Theorem 1 does *not* in any way depend on the generation output of the resources. It is, in fact, equal to the denominator in (3). On the other hand, given the negative signs of the other three terms, they all act as *penalty factors* in calculating the performance score. The first penalty term is impacted by the generation output of the resource only through set Ψ. The second penalty term is impacted by the generation output of the resource through not only set Φ but also the amount of the absolute regulation errors inside the summation. Note that, if $\tau\in\Phi$, then this penalty term is counted twice for time slot $[\tau-1]$. The last term is the absolute regulation error at the terminal time slot $\tau=T$. We can see that the penalty terms for any $\tau\le T$ may be *double counted* if it belongs to sets Ψ and Φ. They may also be simply *ignored* if they belong to set Ω for all $\tau<T$.

In Case 2, we have $\Phi=\{\}$, $\Psi=\{\}$, $\Omega=\{1,2,3,4,5\}$ for Resources 1, and 2. Thus, the performance of the resources at time slots 1-4 are *not* considered by MISO. Here, MISO *ignores* the better performance of Resource 2 at time slots 2-4. But it *does* consider the better performance of Resource 1 at the time slot 5. Thus, MISO drastically favors Resource 1.

In Case 3, $\Phi=\{5\}$, $\Psi=\{3,4\}$, $\Omega=\{1,2\}$ for both resources. The performance of the resources at time slots 1 and 2 are not considered by MISO. Further, MISO *ignores* the errors of the resources at time slots 3 and 4. Instead it penalizes them based on the penalty factor of mileage, i.e., the first penalty factor in (4). Finally, the error at time slot 4 is *double counted*; because $\Phi=\{5\}$. Thus, the better performance of Resource 1 at time slots 2 and 3 is neglected and the better performance of Resource 2 at time slot 4 is counted twice.

V. Recommended Model Refinements

It appears that a "good resource" is defined by CAISO as one that shows low absolute errors in following AGC setpoints at *individual* time slots; and by MISO as one that moves towards the AGC setpoints across *consecutive* time slots. It is hard to argue which approach is better. Nevertheless, it is observed that some very small refinements may improve each metric. For example, we may refine the metric of CAISO as:

$$\mathrm{PS}_{\mathrm{CAISO}}^{\mathrm{Revised}} = \left[1-\frac{\sum_{\tau=1}^{T}\mid s[\tau]-y[\tau]\mid}{\sum_{\tau=1}^{T}|s[\tau]-y[0]|}\right]^{+}, \tag{7}$$

where we simply replace $s[\tau]$ inside the summation in the denominator in (2) by $|s[\tau]-y[0]|$, somewhat similar to in (3). The revised metric is no longer too sensitive to the magnitude of the AGC signal; thus, the issue in Case 4 is resolved, where $\mathrm{PS}_{\mathrm{CAISO}}$ changes to 0.34 for *both* Resources 1 and 2.

As for the concerns with MISO's performance score in Case 2, the term inside the summation in the denominator in (3) can be replaced with $|s[\tau]-y[\tau-1]|$, as follows:

$$\mathrm{PS}_{\mathrm{MISO}}^{\mathrm{Revised}} = \left[1-\frac{\sum_{\tau=1}^{T}\mid s[\tau]-y[\tau]\mid}{\sum_{\tau=1}^{T}\mid s[\tau]-y[\tau-1]\mid}\right]^{+}. \tag{8}$$

In such case, the ideal movement is *recalculated* at each four seconds interval. As a result, the ideal movement is no longer totally independent of the resource movement at each four seconds. $\mathrm{PS}_{\mathrm{MISO}}$ now changes from 1 to 0.25 in Case 2.

VI. Conclusions and Future Work

The different approaches taken by CAISO and MISO to set forth their performance scores is a fact stemming from CAISO and MISO being different markets, with different levels of demand for regulation, etc. Nevertheless, it is important to understand the different implications of these two different practical approaches and the root causes for such implications. This open problem was addressed in this paper. We showed that at least some of the issues that exist in practice, as raised by ISOs, come directly and provably from the core formulations of the performance metrics used by the ISOs. Thus, some refinements are recommended on each metric. Of course, it is ultimately up to the ISOs to decide whether to keep or refine their existing methods or adopt a new method. The mathematical models developed in this paper for performance scoring methods can be used also for broader analysis of the CAISO and MISO markets; such as to study how different resources may strategically respond to these scoring methods; as well as the impact on the overall market effectiveness under various physical and virtual bidding scenarios.

References

[1] A. D. Papalexopoulos and P. Andrianesis, "Performance-based pricing of frequency regulation in electricity markets," *IEEE Trans. on Power Systems*, vol. 29, no. 1, pp. 441–449, Dec. 2014.

[2] A. Sadeghi-Mobarakeh and H. Mohsenian-Rad, "Optimal bidding in performance-based regulation markets: An mpec analysis with system dynamics," *IEEE Trans. on Power Systems*, vol. 32, no. 2, pp. 1282–1292, 2017.

[3] California ISO. [Online]. Available: https://www.caiso.com/Documents/Jul31_2014_Order755MarketDesignReport_ER12-1630_ER14-971.pdf

[4] ——, "Business Requirements Specification - Pay For Performance Regulation," May 2013.

[5] Y. Chen, R. Leonard, M. Keyser, and J. Gardner, "Development of performance-based two-part regulating reserve compensation on miso energy and ancillary service market," *IEEE Trans. on Power Systems*, vol. 30, no. 1, pp. 142–155, May 2015.

[6] www.caiso.com/Documents/PayForPerformanceMileageExample.xls.

[7] A.-H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. on Smart Grid*, vol. 1, no. 3, pp. 320–331, 2010.