Development of Paddy Yield Gap Between Java and Outside Java: Does It Have a Contribution to Paddy Yield Improvement from 2018 to 2021?

Kadir, Kadir and Prasetyo, Octavia Rizky

BPS-Statistics Indonesia, BPS-Statistics Indonesia

1 August 2023

Development of Paddy Yield Gap Between Java and Outside Java: Does It Have a Contribution to Paddy Yield Improvement from 2018 to 2021?

Kadir[1], Octavia Rizky Prasetyo[1]

[1]*BPS-Statistics Indonesia*

Abstract. Increasing the paddy yield is crucial for Indonesia to maintain its national rice sufficiency amid the consistent depletion of wetland paddy areas. In this regard, the yield disparities between regions are challenging, particularly between Java and outside Java. Our study aims to examine the development of the paddy yield gap between the two regions from 2018 to 2021 and its contribution to paddy yield improvement during the period. Using the results of the National Crop-cutting Survey, we found that while the paddy yield in Java outperformed the paddy yield outside Java, the yield difference between the two regions narrowed from around 26 per cent in 2018 to 22 per cent in 2021 due to the increase of the yield outside Java. The results of the Blinder-Oaxaca decomposition suggested that the narrowing gap has a significant contribution to the national paddy yield increase from 2018 to 2021. Our finding confirms that narrowing the yield gap between the two regions by increasing the yield outside Java is crucial to improving paddy yield in Indonesia. Our study also pointed out that improvement in irrigation systems, fertilizer use, and fertilizer assistance are important factors in maintaining the paddy yield and narrowing the gap.

Keywords*: yield, gap, paddy, Blinder-Oaxaca, decomposition*

1. Introduction

Rice is still the main staple food for most Indonesians, by which the rice consumption per capita was around 112 kilograms in 2019 [1]. That high rice consumption, coupled with the growing population, increases rice demand every year. This is a big challenge for Indonesia since at the same time the areas of wetland paddy and agricultural land as a whole are consistently depleting due to economic expansion and transformation of the country from an agriculture-based economy to a manufacturing and services economy [2]. To deal with those circumstances, boosting the national paddy yield becomes inevitable. In this regard, Indonesia faces the challenge of an extensive yield disparity between regions, particularly between Java and outside Java.

Paddy yield is the outcome of the interaction between social and ecological factors. Ecologically, it is impacted by environmental factors, either biotic (climate, soil fertility, water availability) or abiotic such as pest attacks [3]. Socially, it is influenced by farmers' decisions on farm management like input uses, pest attack management, the type of variety choice, technology adoption, etc. [4]. In this regard, farmers' sociodemographic characteristics may influence decisions made related to the production system.

Indonesia is quite massive geographically. It brings consequences of wide ecological differences among regions, particularly between Java and outside Java. Regarding agricultural cultivation, the difference in soil fertility and climate makes the productivity gap between the two regions inevitable. Ecologically, Java Island is blessed with a better climate and more fertile soil compared to outside Java [5][6]. Indeed, on average the paddy yield in Java is higher than outside Java [7]. It is both a challenge and an opportunity for Indonesia to boost its paddy yield.

Increasing the paddy yield outside Java which is lagging is also a matter of the country's food security. Although Java has a higher yield, it has to feed a larger population since most Indonesians are concentrated on the island [8]. Therefore, relying only on the island could not guarantee national rice sufficiency in the future. Therefore, narrowing the yield gap between Java and outside Java by improving the yield outside Java is critical to pursuing the country's food security. However, it needs knowledge of the development of the gap from time to time and its contribution to the yield change over time. In this regard, to the best of our knowledge, studies focusing on the development of the paddy yield gap between the two regions over time that are based on a national survey to provide a national inference and representation are not yet available in the Indonesian context. Our study aims to address this gap by examining the development of the paddy yield gap between the two regions from 2018 to 2021 and its contributing factors. We exercised the results of a nationwide crop-cutting survey called *Survei Ubinan* conducted by Statistics Indonesia on a regular basis.

2. Methodology

Yield gap and its decomposition

Following the literature review before, in our study, we assume that the paddy yield is influenced by a set of variables that reflects the cultivation system's characteristics run by farmers and some control variables, i.e. type of land for paddy cultivation, fertilizer use, water sufficiency, pest attack intensity, climate change impacts, government's fertilizer assistance, and farmers group membership. Therefore, the regression model to estimate the gap is formulated as follows

$$y_i = \gamma ISLAND_i + \boldsymbol{X}_i' \boldsymbol{\beta} + \varepsilon_i \quad (1)$$

In Equation (1), y_i is the natural logarithmic of paddy yield cultivated by *i-th* farmer, $\boldsymbol{X}_i'$ is a vector of variable related to farmers' cultivation characteristics, i.e. type of paddy variety, fertilizer use, water sufficiency, pest attack intensity, climate change impacts, government's fertilizer assistance, and farmers group membership; $\boldsymbol{\beta}$ is a vector of regression coefficients for each variable in the model; ε_i is error components assumed to follow classic linear regression assumptions.

From Equation (1), the difference (gap) of the yield between Java and outside Java is obtained by estimating γ, using the Ordinary Least Square (OLS) method, which shows how much the paddy yield outside Java is smaller than Java in a percentage point. $ISLAND_i$ is the region of *i-th* farmer coded 1 if a farmer is outside Java and 0 if the farmer resides in Java (reference category). More precisely, the yield gap between the two regions was estimated using the formula of $\left(e^{\hat{\gamma}} - 1\right) \times 100\%$. We estimated Equation (1) for the years 2018 and 2021 separately to see the development of the paddy yield gap from 2018 to 2021. To examine the contributing factors in explaining the development of the gap from 2018 to 2021, we applied Blinder-Oaxaca's [9][10] decomposition. The decomposition was started by estimating the paddy yield as the function of explanatory variables for each region of Java and outside Java on this equation:

$$y_i^k = X_i'^k \boldsymbol{\beta}_k + \varepsilon_i, k \in \{0,1\} \quad (2)$$

Estimation of Equation (2) is conditional to the value assigned to k representing the region of farmers (coded 0 for Java and 1 four outside Java). Therefore, the yield gap between Java and outside Java can be estimated as follows:

$$\Delta\bar{y} = \bar{y}^0 - \bar{y}^1 = \bar{X}_i'^0 \widehat{\boldsymbol{\beta}}_0 - \bar{X}_i'^1 \widehat{\boldsymbol{\beta}}_1 = \left[\bar{X}'^0_i \widehat{\boldsymbol{\beta}}_0 - \bar{X}_i'^1 \widehat{\boldsymbol{\beta}}_0\right] + \left[\bar{X}_i'^1 \widehat{\boldsymbol{\beta}}_0 - \bar{X}_i'^1 \widehat{\boldsymbol{\beta}}_1\right] \quad (3)$$

The $\bar{X}_i^1 \widehat{\boldsymbol{\beta}}_0$ component estimates the contrafactual yield of farmers outside Java when they have the yield level equal to farmers in Java due to having the same characteristics as farmers in Java. In this study, we treat farmers in Java as the reference categories, by which their yield is assumed to always be larger than farmers outside Java although the two groups have the same characteristics ($\boldsymbol{\beta}_0 = \boldsymbol{\beta}_* > \boldsymbol{\beta}_1$). In this regard, Java farmers' coefficients are treated as the reference coefficients. The further modification of Equation (3) gives us:

$$\Delta\bar{y} = \left(\bar{X}'^0_i - \bar{X}_i'^1\right)\widehat{\boldsymbol{\beta}}_0 + \bar{X}'^1_i\left(\widehat{\boldsymbol{\beta}}_0 - \widehat{\boldsymbol{\beta}}_1\right) \quad (4)$$

In Equation (4), $\left(\bar{X}'^{0}_{i} - \bar{X}'^{1}_{i}\right)\widehat{\boldsymbol{\beta}}_0$ represents gap component sourced from the difference in characteristics between farmers in Java and outside Java (the explained part) while $\bar{X}'^{1}_{i}\left(\widehat{\boldsymbol{\beta}}_0 - \widehat{\boldsymbol{\beta}}_1\right)$ represents the unobservable gap component sourced from the difference in the coefficients (the unexplained part). It contains the contribution of other variables that are not specified in the model or are unobservable, such as the difference in famers' socio-demographic characteristics, soil fertility, and climate conditions that affect the difference in yield between the two regions. The issue in Equation (4) is there is no standard for determining the reference group or known as the index problem [11]. To address the issue, we also estimate Equation (4) with other difference scenarios of reference coefficients, i.e. $\boldsymbol{\beta}_* = (n_0\widehat{\boldsymbol{\beta}}_0 + n_1\widehat{\boldsymbol{\beta}}_1)/(n_0 + n_1)$, where *n* is the number of observations or farmers in each group [12], $\boldsymbol{\beta}_* = 0.5$ or equal weight [13], and $\boldsymbol{\beta}_*$ equal to the estimation of pooled regression model in Equation (2) or pooled coefficients [14].

To examine the contribution of the development of the yield gap from 2018 to 2021 to the change in the yield from 2018 to 2021, we added a time dummy variable in Equation (1) coded 1 if the observation was in 2018 and 0 if the observation was in 2021 (reference category). Therefore, the regression model can be represented as follows:

$$y_i = \gamma ISLAND_i + \delta YEAR_i + \boldsymbol{X}'_i\boldsymbol{\beta} + \varepsilon_i \quad (5)$$

In Equation (5), $\left(e^{\hat{\delta}} - 1\right)$ denotes how much the paddy yield in 2018 is lower than the yield in 2021. The expected sign for $\hat{\delta}$ is negative showing that paddy yield increased from 2018 to 2021. We also estimate Equation (5) using quantile regression to determine the difference of the yield around the distribution. In this regard, following [15], we assume that $Q_{Y|X}(\tau) = min\,\{\eta\,|P(y \leq \eta|x) \geq \tau\}, \forall\tau \in (0,1)$ and $Q_{Y|X}(\tau) = \boldsymbol{X}'_i\boldsymbol{\beta}(\tau)$, where *X* represents all independent variables in Equation (5) and τ the quantile of paddy yield distribution. We then decomposed the yield gap at the mean between 2021 and 2018 using Blinder-Oaxaca's decomposition by treating 2021 as the reference category. By doing so, it could be confirmed that the yield gap between Java and outside Java has a significant contribution to the paddy yield development from 2018 to 2021. To ensure the validity of inference under the violation of the homoscedasticity assumption for the regression residuals, we applied a robust covariance estimation for OLS [16] and quantile regressions [17].

Data and source of data

Our study uses the microdata of the results of a nationwide survey called the Crop-cutting Survey or *Survey Ubinan* conducted by BPS on a regular basis throughout the year. The survey is performed in all provinces and is dedicated to collecting information on the yield of the food crops, including paddy, and other variables related to the yield. In this study, we made use of the survey results of the years 2018 and 2021 consisting of 68,440 and 56,885 samples of paddy households respectively. The explanation for all variables used in the analysis obtained from the results of the survey is presented in Table 1. All variables are based on farmers' declarations during the interview.

Table 1. Description of the variables

Dependent variables:	Natural logarithmic of the yield in the years 2018 and 2021
Independent variables:	
Region	The region consists of two categories: Java coded 0 (the reference categories) and outside Java coded 1
Type of land for cultivation	It is divided into three categories: dryland coded 0 (the reference category), wetland without an irrigation system coded 1, and wetland with an irrigation system coded 2.
Fertilizer use	The variable is divided into two categories: not using fertilizer coded 0 (the reference category) and using fertilizer coded 1.
Water sufficiency	The variable represents water availability for paddy cultivation in the current year compared to the previous year. It is divided into two categories: the water

Dependent variables:	Natural logarithmic of the yield in the years 2018 and 2021
Independent variables:	
	sufficiency is sufficient compared to the previous year coded 0 (the reference category) and the water sufficiency is not sufficient compared to the previous year coded 1.
Pest attack intensity	The variable represents the intensity of pest attacks in the current year when enumeration was performed compared to the previous year. It is divided into two categories: the intensity decreases or is still the same compared to the previous year coded 0 (the reference category) and the intensity increases coded 1.
Climate change impact	The variable reflects farmers' experience of the climate change impacts in the form of drought or/and flood. It is divided into two categories: not experiencing the impact of the climate change coded 0 (the reference category) and experiencing the climate change coded 1.
Government's fertilizer assistance	The variable represents the fertilizer assistance received by the farmers from the government both central and regional governments. It is divided into two categories: not receiving the assistance coded 0 (the reference category) and receiving the assistance coded 1.
Farmers group membership	The variable represents the farmers' membership in the farmers' group. It is divided into two categories: not farmers group membership coded 0 (the reference category) and farmers group membership coded 1.

3. Results and discussion

Cultivation characteristics

Most of the paddy household samples are outside Java both in the years 2018 and 2021. However, this figure does not reflect the real distribution of paddy households in Java and outside Java since the sample weight is not available from the survey to be used in the estimation. Moreover, the Crop-cutting Survey for paddy makes use of the results of the Area Sampling Frame (ASF) Survey as the sampling frame. Therefore, the distribution of the samples follows the distribution of the paddy area, in which the paddy area outside Java is higher than in Java [1]. The number of paddy households or farmers in Java is larger than outside Java [18]. Concerning the type of land for paddy cultivation, most farmers (paddy households) samples cultivated their paddy crops in the wetland with an irrigation system both in 2018 (59.68 percent) and 2021 (57.39 percent). However, the relatively high proportion of samples cultivating their paddy crops in wetlands without an irrigation system, which was around 32 percent in 2018 and 2021, should get serious attention from the government in boosting the paddy yield since an irrigation system is very crucial in maintaining the paddy yield. The improvement in the irrigation system must be focused outside of Java since the proportion of irrigated wetlands in this region is lower than in Java.

Most of the samples used fertilizer for their paddy cultivation in 2018 and 2021 which are 94.34 percent and 95.41 respectively. However, the proportion of samples receiving fertilizer assistance from the government was only 59.17 percent in 2018 and 68.22 percent in 2021. When it comes to climate change impacts in the form of droughts and/or floods, only 17.46 percent of household samples declared experiencing climate change impacts in 2018 and 16.79 percent in 2021. However, those proportions are alarming and call for the need for climate-smart agriculture as recommended by the Food and Agriculture Organization (FAO) to mitigate the impact of climate change on paddy production [19].

The proportion of samples declaring there was an increase in pest attack intensity compared to the previous year is quite high, which is 76.87 percent in 2018 and 81.10 percent in 2021. It is a critical issue since a high intensity of pest attacks possibly could lower the paddy yield. With regard to water sufficiency, most of the samples stated that water availability was sufficient compared to the previous year in 2018 (83.89 percent) and in 2021 (91.80 percent).

Most of the samples stated that they were members of farmers' groups during the interview both in 2018 (68.99 percent) and 2021 (71.18 percent). It is a good sign since farmers' group membership is very important for farmers to have access to government assistance to increase their yield. It could be explained since in practice government assistance is distributed through the farmers' groups. Besides, the existence of farmers' groups is very critical for farmers to gain knowledge and best practices regarding paddy cultivation. Moreover, the paddy yield in the form of natural logarithmic in 2021 was higher than the paddy yield in 2018. It confirmed that there was a paddy yield increase during the period of study.

In terms of cultivation characteristics, Java outperformed outside Java in irrigation systems, fertilizer use, and fertilizer government assistance. The proportion of samples cultivated paddy crops in irrigated wetlands in Java is higher than outside Java both in 2018 dan 2021. Likewise, the proportion of samples using fertilizer and receiving fertilizer assistance from the government was also higher in Java than outside Java in both 2018 and 2021. Moreover, the paddy yield in the natural logarithmic term in Java was quite higher than outside Java both in 2018 and 2021. It anticipates that the estimation of the paddy yield gap between the two regions will be quite substantial in 2018 and 2021.

Table 2. Summary of the variables

Variables	2018			2021		
	Java	Outside Java	Total	Java	Outside Java	Total
Continuous variable (mean)						
ln(paddy yield)	1.89	1.58	1.72	1.88	1.63	1.74
Categorical variable (%)						
1. Type of land cultivated						
Dryland	7.24	8.75	8.06	10.47	10.51	10.49
Not irrigated wetland	24.98	38.49	32.27	23.58	39.09	32.12
Irrigated wetland	67.78	52.75	59.68	65.95	50.40	57.39
2. Fertilizer use						
Not use fertilizer	0.54	10.04	5.66	0.64	7.82	4.59
Use fertilizer	99.46	89.96	94.34	99.36	92.18	95.41
3. Fertilizer assistance						
Not received	34.43	46.29	40.83	23.10	38.87	31.78
Received	65.57	53.71	59.17	76.90	61.13	68.22
4. Farmer group						
Not member	33.10	29.23	31.01	34.24	24.40	28.82
Member	66.90	70.77	68.99	65.76	75.60	71.18
5. Climate change impact						
Not experienced	87.11	78.63	82.54	88.41	78.96	83.21
Experienced	12.89	21.37	17.46	11.59	21.04	16.79
6. Increase in pest attack						
Not experienced	21.85	24.22	23.13	16.82	20.61	18.90
Experienced	78.15	75.78	76.87	83.18	79.39	81.10
7. Water insufficiency						
Not experienced	84.11	83.70	83.89	92.16	91.51	91.80
Experienced	15.89	16.30	16.11	7.84	8.49	8.20
8. Region	46.06	53.94	100.00	44.97	55.03	100.00

Note: *The total number of samples is 125,325 paddy households*

Development of the yield gap and its decomposition

The OLS estimation in Table 3 shows that the paddy yield gap between Java and outside Java was narrowed from 2018 to 2021 although the paddy yield outside Java is still lower than Java. It pointed out that paddy yield outside Java get closer to Java during the period. It is also confirmed by the kernel density distribution of the yield of the two regions in 2018 and 2021 presented in Figure 1. The paddy yield distribution outside Java getting closer to the distribution of the paddy yield in Java from 2018 to 2021.

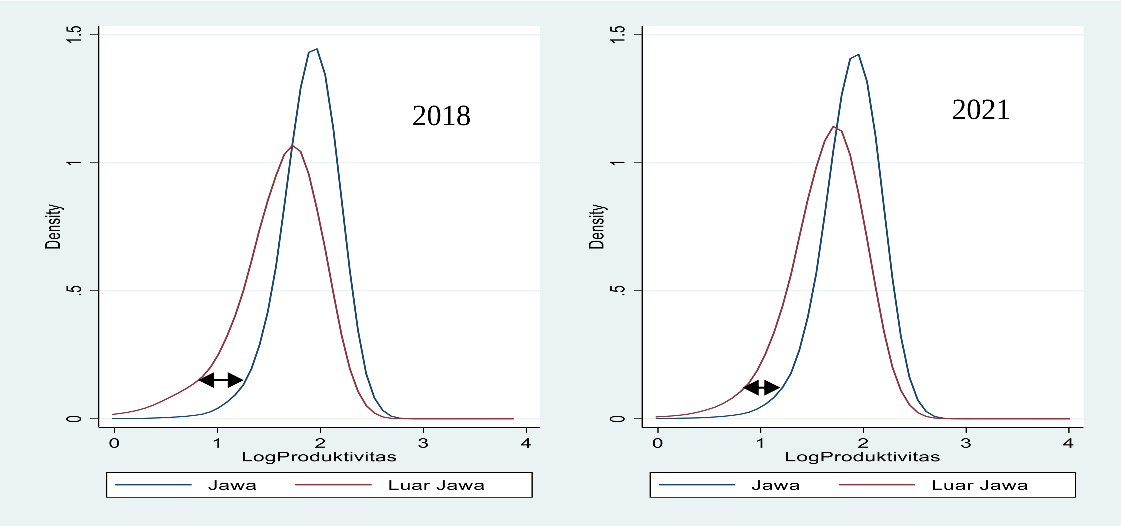

Figure 1. Development of paddy yield distribution by region, 2018-2021[1]

Although narrowing, the yield gap between the two regions is still substantial. Without controlling other variables, in 2018, the paddy yield outside Java was 26 percent ($(e^{-0.3076} - 1) \times 100\%$) lower than the paddy yield in Java. In 2021, the paddy yield outside Java was 22 percent lower than the yield in Java. In other words, the paddy yield gap between the two regions narrowed from 2018 to 2021. However, after controlling other variables that also explain the yield, the gap dropped to 21 percent in 2018 and 18 percent in 2021. Moreover, the quite substantial yield gap between the two regions shows that addressing the gap by improving the paddy yield outside Java is still a challenge that must be addressed.

Table 3. OLS estimation of the yield gap equation

Dependent variable: ln(paddy yield)	OLS estimation 2018		OLS estimation 2021	
	Base	Full model	Base	Full model
Region				
Outside Java	-0.3076***	-0.2412***	-0.2444***	-0.1966***
	(0.0028)	(0.0025)	(0.0029)	(0.0029)
Type of land cultivated				
Not irrigated wetland		0.2840***		0.0011
		(0.0067)		(0.0060)
Irrigated wetland		0.4051***		0.1075***
		(0.0064)		(0.0055)
Fertilizer use				
Use fertilizer		0.3308***		0.2075***
		(0.0083)		(0.0100)
Fertilizer assistance				
Received		0.0559***		0.0560***
		(0.0028)		(0.0036)

[1] We generated the kernel distribution using the Epanechkinov kernel function. The bandwidth is 0.15.

Dependent variable: ln(paddy yield)	OLS estimation 2018		OLS estimation 2021	
	Base	Full model	Base	Full model
Farmer group membership				
Member		0.0457***		0.0132***
		(0.0028)		(0.0032)
Climate change impact				
Experienced		-0.0935***		-0.0985***
		(0.0045)		(0.0047)
Increase in pest attack intensity				
Experienced		-0.0255***		-0.0323***
		(0.0030)		(0.0036)
Water insufficiency				
Experienced		-0.0895***		-0.1171***
		(0.0045)		(0.0062)
Constant	1.8860***	1.1905***	1.8723***	1.5907***
	(0.0016)	(0.0092)	(0.0018)	(0.0121)
R-squared	0.1408	0.3346	0.1056	0.1848

Note: *The number of observations is 125,325 paddy households. *** is significant at a 1 percent level of significance. Robust standard errors to heteroscedasticity are presented in parentheses.*

As expected, the type of land cultivated for paddy crops has a significant impact on the paddy yield. Compared to the dray-land as the reference category, paddy crops cultivated in a wetland without an irrigation system have a 33 percent higher yield while paddy yields cultivated in a wetland with an irrigation system had a 50 percent higher yield in 2018. However, in 2021 only paddy crops cultivated in irrigated wetlands had significantly higher yields than those cultivated in drylands. It confirms the critical role of irrigation systems in boosting yield.

Our study also confirms the critical role of fertilizer in maintaining the yield. Farmers using fertilizer significantly had larger yields, which were 34 percent higher in 2018 and 23 percent higher in 2021 compared to farmers who did not use fertilizer in their paddy cultivation. Our study is supported by [20] findings. Therefore, the availability of fertilizer when needed by farmers is very important. In this regard, our study pointed out that fertilizer assistance from the government is also crucial in maintaining the paddy yield. The estimation results showed that farmers receiving fertilizer assistance from the government significantly had a higher yield in 2018 and 2021 compared to their counterparts who did not receive fertilizer assistance from the government.

The estimation results also pointed out that almost all controlling variables are statistically significant at a 5 percent level of significance in explaining the variation of the yield among farmers both in 2018 and 2021. As expected climate change has a negative impact on the paddy yield. Despite based on only the farmers' declaration, the estimation results pointed out that farmers experiencing climate change impacts in the form of drought or floods had significantly lower yields compared to those who did not experience climate change impacts, which were 9 percent lower in 2018 and 2021. In line with the impact of climate change on the yield, farmers experiencing higher intensity of pest attacks compared to the previous year had significantly lower yields compared to their counterparts who did not experience an increase in pest attacks. The results are consistent for both 2018 and 2021. Moreover, water sufficiency is also pivotal in maintaining the yield. Based on the estimation results, farmers experiencing water insufficiency in 2018 and 2021 had significantly lower yields than those who did not experience water scarcity. It is also important to note that a high pest attack intensity and water scarcity may have an association with climate factors [21]. Therefore, the impact of climate change on the yield is substantial and should be taken considerably in agricultural policy.

Our study also confirms the importance of farmers' group membership in maintaining the paddy yield, where farmers who were members of the farmers' group significantly had a higher yield than those who were not members in both 2018 and 2021. Our study is also consistent with [20]. It could be explained since by being a member of a farmer group a farmer will have access to any means to maintain and increase the yield from knowledge, best practices, to government assistance.

The results of decomposition using Blinder-Oaxaca pointed out that the difference in cultivation characteristics only explained around 15 percent (0.0460/0.3076) to 22 percent (0.0664/0.3076) of the total yield gap in 2018. The main contributors in explaining the paddy yield gap between the two regions are the type of land for cultivation and fertiliser uses, which explained the gap of 7 percent and around 4-10 percent respectively. Moreover, the results of gap decomposition for the year 2021 pointed out that the difference in cultivation characteristics and other control variables could explain around 15-20 percent of the gap while the rest 85-80 percent is unexplained. Like in 2018, among variables that explain the gap, the type of land is the main contributor to the gap with a contribution of about 6-7 percent in explaining the total gap. It means focusing on the improvement of the variable outside Java could have a significant impact on lowering the gap in the future. Indeed, although outside Java performed a significant improvement for the betterment of the variable, as pointed out in Table 2, the region is still lagging behind its Java counterparts. It implies that the improvement on the variable outside Java must continue to be pursued in order to boost the region's paddy yield and narrow the yield gap.

The decomposition results also confirmed a significant contribution of climate change impact in explaining the paddy yield gap between the two regions. The contribution even increased from 2-3 percent in 2018 to 3-4 percent in 2021. The finding confirmed the importance of climate change impact mitigation in narrowing the productivity gap between the two regions. However, using the same data for 2021, [20] pointed out that the extent of the influence of the climate change impact on lowering the yield is higher in Java than outside Java. Therefore, the mitigation must be implemented comprehensively both in Java and outside Java.

Table 4. Decomposition results of the yield gap between Java and outside Java

	2018		2021	
	Proportional coefficient	Pooled coefficient	Proportional coefficient	Pooled coefficient
Total gap	0.3076***		0.2444***	
	(0.0028)		(0.0029)	
Explained gap				
Type of land cultivated	0.0218***	0.0225***	0.0157***	0.0165***
	(0.0010)	(0.0010)	(0.0007)	(0.0007)
Fertilizer use	0.0129***	0.0314***	0.0069***	0.0149***
	(0.0010)	(0.0010)	(0.0009)	(0.0008)
Fertilizer assistance	0.0062***	0.0066***	0.0081***	0.0088***
	(0.0004)	(0.0004)	(0.0006)	(0.0006)
Farmer group membership	-0.0017***	-0.0018***	-0.0012***	-0.0013***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Climate change impact	0.0071***	0.0079***	0.0082***	0.0093***
	(0.0004)	(0.0005)	(0.0005)	(0.0005)
Increase in pest attack intensity	-0.0007***	-0.0006***	-0.0012***	-0.0012***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Water insufficiency	0.0004	0.0004	0.0008***	0.0008***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Total	0.0460***	0.0664***	0.0373***	0.0478***
	(0.0016)	(0.0016)	(0.0015)	(0.0014)
Unexplained gap				
Type of land cultivated	-0.1336***	-0.1343***	-0.0663***	-0.0672***
	(0.0114)	(0.0113)	(0.0094)	(0.0094)
Fertilizer use	-0.7200***	-0.7386***	-0.4865***	-0.4945***
	(0.0422)	(0.0430)	(0.0545)	(0.0552)
Fertilizer assistance	-0.1132***	-0.1137***	-0.1167***	-0.1174**
	(0.0086)	(0.0086)	(0.0117)	(0.0117)

	2018		2021	
	Proportional coefficient	Pooled coefficient	Proportional coefficient	Pooled coefficient
Farmer group membership	0.0802***	0.0802***	0.0600***	0.0601***
	(0.0094)	(0.0094)	(0.0110)	(0.0110)
Climate change impact	0.0464***	0.0456***	0.0714***	0.0702***
	(0.0097)	(0.0096)	(0.0104)	(0.0103)
Increase in pest attack intensity	-0.0211**	-0.0212**	0.0036	0.0036
	(0.0102)	(0.0102)	(0.0127)	(0.0127)
Water insufficiency	-0.0433***	-0.0433***	-0.0754***	-0.0754***
	(0.0100)	(0.0100)	(0.0131)	(0.0131)
Constant	1.1663***	1.1663***	0.8172***	0.8171***
	(0.0457)	(0.0457)	(0.0593)	(0.0593)
Total	0.2616***	0.2412***	0.2071***	0.1966***
	(0.0027)	(0.0025)	(0.0030)	(0.0028)

Note: *The number of observations is 125,325 paddy households. *** is statistically significant at 1 percent level of significance. Robust standard errors to heteroscedasticity are presented in parentheses.*

Contribution of the narrowing gap to the yield increase

The estimation results of the regression model with the year dummy variable (Equation 5) pointed out that the paddy yield in 2021 is significantly higher than the yield in 2018. In other words, there is a significant increase in the yield from 2018 to 2021. Without controlling other variables influencing the paddy yield, the estimation results pointed out that the paddy yield in 2018 was 2 percent lower than the yield in 2021. However, after controlling other variables impacting the yield, the difference narrowed to only less than one percent. It shows that the increase in the paddy yield from 2018 to 2021 was relatively small and can be explained by the development of dependent variables during the period. Moreover, the estimation results of quantile regression pointed out that the higher increase of the paddy yield during the period happens on the upper part of the yield distribution, which is at the 95th quantile. In line with that, the highest yield gap between Java and outside occurs on the lower part of the yield distribution, which is the 25th quantile. It confirms that there is a significant opposite association between the yield gap between the two regions and the yield difference between 2018 and 2021. In other words, the narrowing yield gap between the two regions may have a significant contribution to the paddy yield increase from 2018 to 2021.

Table 5. OLS estimation of the yield difference using OLS and quantile regression

Dependent variable: ln(paddy yield)	OLS estimation		Quantile regression		
	Base	Full model	Q_{15}	Q_{55}	Q_{95}
Yield difference					
2018	-0.0176***	-0.0099***	0.2002***	0.0063***	-0.0130***
	(0.0022)	(0.0019)	(0.0031)	(0.0019)	(0.0027)
Yield gap by region					
Outside Java		-0.2215***	-0.2840***	-0.1907***	-0.1223***
		(0.0019)	(0.0033)	(0.0019)	(0.0026)
Type of land cultivated					
Not irrigated wetland		0.1419***	0.2386***	0.1006***	0.0174***
		(0.0046)	(0.0089)	(0.0051)	(0.0055)
Irrigated wetland		0.2543***	0.3806***	0.2040***	0.0940***
		(0.0043)	(0.0088)	(0.0049)	(0.0052)
Fertilizer use					
Use fertilizer		0.3090***	0.4200***	0.2733***	0.1584***
		(0.0064)	(0.0109)	(0.0062)	(0.0073)
Fertilizer assistance					

Dependent variable: ln(paddy yield)	OLS estimation		Quantile regression		
	Base	Full model	Q_{15}	Q_{55}	Q_{95}
Received		0.0548***	0.0708***	0.0503***	0.0130***
		(0.0022)	(0.0036)	(0.0021)	(0.0028)
Farmer group membership					
Member		0.0362***	0.0288***	0.0427***	0.0594***
		(0.0021)	(0.0033)	(0.0021)	(0.0028)
Climate change impact					
Experienced		-0.0927***	-0.1187***	-0.0745***	-0.0545***
		(0.0033)	(0.0051)	(0.0032)	(0.0041)
Increase in pest attack					
Experienced		-0.0276***	-0.0313***	-0.0255***	-0.0325***
		(0.0023)	(0.0038)	(0.0022)	(0.0031)
Water insufficiency					
Experienced		-0.0973***	-0.1305***	-0.0842***	-0.0580***
		(0.0037)	(0.0059)	(0.0035)	(0.0028)
Constant		1.3560***	0.8875***	1.4692***	2.0086***
		(0.0077)	(0.0126)	(0.0078)	(0.0090)
R-squared		0.2577	0.2556	0.2567	0.2435

Note: *The number of observations is 125,325 paddy households. *** is statistically significant at a 1 percent level of significance, ** is statistically significant at a 5 percent level of significance. Robust standard errors to heteroscedasticity are presented in parentheses.*

Blinder-Oaxaca decomposition results strongly supported the importance of the paddy yield gap narrowing between Java and outside Java to increase the paddy yield. It is reflected by the significance and the sign of estimated coefficients for the variables in the explained part of the decomposition results. The coefficients are significant at a 5 percent level of significance and have negative signs meaning that the difference in the paddy yield between 2018 and 2021 will be higher (or the yield increase will be higher) if the paddy gap between the two regions is eliminated from the regression model. In other words, the narrowing gap between the two regions has a significant contribution to the yield increase from 2018 to 2021. The estimation result also applies to the type of cultivated land variable. The negative sign of the estimated coefficient for the variable shows that eliminating the gap in irrigation system availability between the two regions by focusing on improvement outside Java could have a substantial impact on increasing the paddy yield. Moreover, the negative sign and significant impact of pest attack intensity estimated coefficient pointed out that the increase of sample proportion reporting that they experienced an increase in pest attack intensity from 2018 to 2021 restrained the paddy yield increase during the period. Therefore, pest attack management should be a priority in maintaining and increasing the paddy yield.

The decomposition results also pointed out that improvement in fertilizer use, fertilizer assistance disbursement, and water sufficiency, particularly outside Java has a substantial contribution to the yield difference between 2018 and 2021 or the yield increase during the period. Those variables contribute around 19 percent, 29 percent, and 46 percent respectively in explaining the yield difference between 2018 and 2021. It means that improvement in those variables between 2018 and 2021 has contributed to the yield increase during the period. In general, all dependent variables all together can explain around 44 percent of the yield difference between 2018 and 2021. It indicates that a substantial increase in the yield could be achieved by focusing on the improvement of those variables.

Table 6. Decomposition results of the yield gap difference between 2018 and 2021

	2021 coefficient	Equal coefficients	Proportional coefficient	Pooled coefficient
Total gap		0.0176***		
		(0.0022)		
Explained gap				
Yield gap Java and Outside Java	-0.0021***	-0.0024***	-0.0024***	-0.0024***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Type of land cultivated	-0.0025***	-0.0061***	-0.0064***	-0.0060***
	(0.0003)	(0.0005)	(0.0005)	(0.0005)
Fertilizer use	0.0022***	0.0029***	0.0029***	0.0033***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Fertilizer assistance	0.0051***	0.0051***	0.0051***	0.0050***
	(0.0004)	(0.0003)	(0.0003)	(0.0002)
Farmer group membership	0.0003***	0.0006***	0.0007***	0.0008***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Climate change impact	0.0007***	0.0006***	0.0006***	0.0006***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Increase in pest attack intensity	-0.0014***	-0.0012***	-0.0012***	-0.0012***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Water insufficiency	0.0093***	0.0082***	0.0081***	0.0077***
	(0.0005)	(0.0004)	(0.0004)	(0.0003)
Total	0.0116***	0.0077***	0.0074***	0.0078***
	(0.0011)	(0.0012)	(0.0012)	(0.0012)
Unexplained gap				
Yield gap Java and outside Java	0.0240***	0.0243***	0.0243***	0.0243***
	(0.0021)	(0.0021)	(0.0021)	(0.0021)
Type of land cultivated	-0.2689***	-0.2653***	-0.2650***	-0.2653***
	(0.0077)	(0.0076)	(0.0076)	(0.0076)
Fertilizer use	-0.2396***	-0.2403***	-0.2403***	-0.2407***
	(0.0253)	(0.0254)	(0.0254)	(0.0254)
Fertilizer assistance	0.0001	0.0001	0.0001	0.0002
	(0.0072)	(0.0074)	(0.0074)	(0.0074)
Farmer group membership	-0.0549***	-0.0553***	-0.0553***	-0.0550***
	(0.0073)	(0.0073)	(0.0073)	(0.0073)
Climate change impact	-0.0059	-0.0059	-0.0059	-0.0059
	(0.0076)	(0.0076)	(0.0076)	(0.0076)
Increase in pest attack intensity	-0.0120	-0.0121	-0.0121	-0.0122
	(0.0083)	(0.0083)	(0.0084)	(0.0084)
Water insufficiency	-0.0320***	-0.0309***	-0.0308***	-0.0305***
	(0.0089)	(0.0086)	(0.0086)	(0.0086)
Constant	0.5953***	0.5953***	0.5953***	0.5953***
	(0.0294)	(0.0294)	(0.0294)	(0.0294)
Total	0.0061***	0.0099***	0.0102***	0.0099***
	(0.0021)	(0.0020)	(0.0020)	(0.0019)

Note: *The number of observations is 125,325 paddy households. *** is statistically significant at a 1 percent level of significance. Robust standard errors to heteroscedasticity are presented in parentheses.*

Our research did not analyse and discuss the influence of the difference in the level of land suitability for paddy crops between Java and outside Java in explaining the yield gap between the two regions due to the limitation in data availability. The crop-cutting survey conducted by BPS does not collect that kind of information. Considering the significance of the information in explaining the gap, it could be the limitation of our research that could be addressed by another research.

4. Conclusion

Our study aims to examine the development of the paddy yield gap between Java and outside Java from 2018 to 2021, which is a challenge for Indonesia in boosting its paddy yield and maintaining the country's food security. We also want to figure out whether the development has a significant contribution to the paddy yield improvement during the period. With respect to those objections, we found that the paddy yield gap between the two regions narrowed significantly from 26 percent in 2018 to 22 percent in 2021. Blinder-Oaxaca decomposition results pointed out that during the period the main contributors in explaining the yield gap were the type of land for cultivation, fertilizer use, and government fertilizer assistance. During the period, the improvement in fertilizer uses and fertilizer government assistance outside Java relative to Java seems to have contributed to the narrowing gap between the two regions from 2018 to 2021. It can be concluded that a consistent improvement in irrigation systems, fertilizer uses, and government fertilizer assistance outside Java will have a substantial impact on the yield increase in the region and narrow the gap. Moreover, our study also pointed out a shred of strong evidence that the narrowing paddy yield gap between the two regions has a significant impact on the yield increase from 2018 to 2021. Therefore, to increase the paddy yield, our study recommends consistently improving irrigation systems as well as fertilizer use and assistance outside Java to increase the region's paddy yield and narrow the gap at the same time.

5. References

[1] BPS 2019 *Luas Panen dan Produksi Padi di Indonesia 2019* (Jakarta: BadanPusat Statistik).

[2] Purbiyanti E, Yazid M, & Januarti I 2017 *Konversi Lahan Sawah di Indonesia dan Pengaruhnya Terhadap Kebijakan Harga Pembelian Pemerintah (HPP) Gabah/Beras*. *Jurnal Managemen & Agribisnis.* **14** 3 p 209-217

[3] Ngoune Liliane T & Shelton Charles M 2020 *Factors Affecting Yield of Crops*. *IntechOpen*. doi: 10.5772/intechopen.90672

[4] Nkurunziza L, Watson CA, Öborn I et al. 2020 *Socio-ecological factors determine crop performance in agricultural systems*. *Sci Rep.* **10** 4232.https://doi.org/10.1038/s41598-020-60927-1

[5] Sukarman & Dariah A 2014 *Tanah Andosol di Indonesia: Karakteristik, Potensi, Kendala, dan Pengelolaannya untuk Pertanian*. *Balai Besar Penelitian dan Pengembangan Sumberdaya Lahan Pertanian, Kementan*

[6] Wahyunto & Widiastuti F 2014 *Lahan Sawah Sebagai Pendukung Ketahanan Pangan serta Strategi Pencapaian Kemandirian Pangan*. *Jurnal Sumberdaya Lahan Edisi Khusus.* 17-30

[7] Amalia RR, Prasetyo OR, & Kadir 2019 *Analisis Kesenjangan Produktivitas Padi di Jawa dan Luar Jawa*. *Proceeding Seminar Nasional Official Statistics 2019.* **1** 1 161-170

[8] BPS. (2023). *Proyeksi Penduduk Indonesia 2020-2050* (Badan Pusat Statistik)

[9] Blinder A 1973 *Wage Discrimination: Reduced Form and Structural Estimates*. *Journal of Human Resources.* **8** 436-455

[10] Oaxaca R 1973 *Male-Female Wage Differentials in Urban Labor Markets*. *International Economic Review.* **14** 693-709

[11] Rahimi E & Hashemi Nazari SS 2021 *A detailed explanation and graphical representation of the Blinder-Oaxaca decomposition method with its application in health inequalities*. *Emerg Themes Epidemiol.* **18** 12. https://doi.org/10.1186/s12982-021-00100-9

[12] Cotton J 1988 *On the Decomposition of Wage Differentials*. *The Review of Economics and Statistics.* **70** 2 236-243

[13] Oaxaca RL and Ransom MR 1994 *On discrimination and the decomposition of wage differentials*. *Journal of Econometrics.* **61** 5–21

[14] Neumark D 1988 *Employers' Discriminatory Behavior and the Estimation of Wage Discrimination*. *Journal of Human Resources.* **23** 3 279-295

[15] Koenker R & Bassett G 1978 *Regression quantiles*. *Econometrics.* **46** 33-50

[16] Wooldridge J M 2003 *Introductory Econometrics: A Modern Approach* (2nd ed. New York: Thomson Learning)

[17] Machado JAF, Parente PMDC, and Santos Silva JMC 2013 qreg2: *Stata module to*

perform quantile regression with robust and clustered standard errors. *Statistical Software Components S457369* (Boston College Department of Economics)
[18] BPS 2018 *Hasil Survei Pertanian Antar Sensus (SUTAS) 2018* (Jakarta, BPS).
[19] FAO 2021 *Climate-smart agriculture 2021* (Food and Agriculture Organization).
[20] Prasetyo OR & Kadir 2023 *Assessing the Impact of Jajar Legowo Planting System on Wetland Paddy Productivity and Income of Farmers in Indonesia* (*pre-print*)
[21] Prasetyo OR & Kadir 2023 *Climate Change Impact on Paddy Yield in Indonesia: Farmers' Experience based on the 2021 Crop-Cutting Survey's Results* (*pre-print*)