WHAT IS *PROPER*?

REPUTATIONAL INCENTIVES FOR POLLUTION CONTROL IN INDONESIA

by

PROPER-PROKASIH Team, BAPEDAL, Jakarta*

PRDEI, World Bank, Washington DC**

November 1995

* Nabiel Makarim, Rasio Ridho, Anton Sarjanto, Ahmad Salim, Made Agus Setiawan, Damayanti Ratunanda, Fredika Wawointana and Ridwan Dahlan

** Shakeb Afsah, Benoit Laplante and David Wheeler

Table of Contents

1. INTRODUCTION 1

2. WHY PROPER? 1

2.1 Reputational Incentives 2

2.2 Incentive Regulation 2

2.3 Rating 3

2.4 Uncertainty 4

2.5 Summary: Implications for PROPER 4

3. PARAMETERS OF PROPER - PHASE I 5

3.1 Scope and Coverage: Which Type of Pollution and Sources Should be Included? 5

3.2 Compliance Assessment: Which Water Pollution Regulations? 6

3.3 The 5-Color Performance Categories 6

4. COMPLIANCE ANALYSIS 7

4.1 Defining Compliance With Water Pollution Regulations 8

4.2 Steps in Compliance Assessment 9

4.3 Information Needed for Compliance Analysis 9

4.4 Sources of Information 10

4.5 Measuring Compliance with Discharge Standards 12

4.6 Reliability of Compliance Assessments 13

4.7 Summary of the Methodologies for Testing Reliability 14

5. ASSIGNING COLOR RATINGS TO POLLUTERS 15

5.1 Beyond Compliance: From Blue to Green 16

5.2 From Green to Gold 16

5.3 Incorporation of Air and Hazardous Waste Assessments 16

5.4 Non-Compliance: From Red to Black 17

5.5 The Final Filter: Collective Judgment in BAPEDAL 18

6. SUMMARY AND CONCLUSIONS 18

8. REFERENCES 20

1. INTRODUCTION

In June 1995, Indonesia launched an innovative program for public disclosure of polluters' environmental performance. This initiative, called the Program for Pollution Control, Evaluation and Rating (PROPER), is expected to serve two objectives:

- Promote compliance with existing regulations
- Reward firms whose performance exceeds regulatory standards

Under PROPER, a polluter is assigned one of five color ratings:

Table 1

Rating	Performance Level
Gold	Excellent
Green	Good
Blue	Adequate
Red	Poor
Black	Very Poor

Although existing programs such as the U.S. Toxic Release Inventory (TRI) are based on public disclosure of pollution data, PROPER is, to our knowledge, the first government program to publish a single index of environmental performance. Indonesia's National Pollution Control Agency (BAPEDAL) is systematically developing and testing the program in collaboration with a team from the World Bank's Policy Research Department (PRDEI) and Country Department EA3. Phase I implementation has focused on 187 polluting facilities scattered across the islands of Java, Sumatra and Kalimantan. It focuses on water pollution, because an appropriate database has already been produced by two BAPEDAL programs: PROKASIH (Clean Rivers) and JAGATIRTA (water regulation enforcement). Starting next year, PROPER is expected to expand toward coverage of air, water and hazardous solid pollution from all medium/large industrial sources in Indonesia. The information and compliance management system developed for the implementation of PROPER will also facilitate and improve the implementation of formal regulations. This same system will also facilitate the introduction of economic instruments (e.g. pollution charges) if it is so desired.

In this paper, we describe Phase I of PROPER: its rationale, methodology and implementation problems. Future work will focus on impact assessment.

2. WHY PROPER?

Enforcement of formal regulation in Indonesia is currently weak, and the modest size of BAPEDAL's budget assures that this weakness will persist in the near future. However,

manufacturing is growing at over 10% annually, and the Indonesian Government recognizes the mounting risk of severe pollution damage. Under these conditions, the Environment Ministry has decided that a large-scale public disclosure program may induce significant pollution abatement while the formal regulatory system is further developed. In this section, we introduce the empirical and conceptual foundations of this approach.

2.1 Reputational Incentives

Who cares what the public knows about pollution? Recent evidence suggests that many firms do indeed care -- in both the OECD and developing countries.[1] Environmental reputation matters for firms whose expected costs or revenues are affected by judgments of environmental performance by customers, suppliers, and stockholders. Many factors can affect Indonesian firms' evaluation of their environmental reputation, including company size, export orientation, and multinational ownership. For reputationally-sensitive companies, public certification of good or bad performance may translate to large expected gains or losses over time.

While there is no 'science' of reputation-formation, it is commonly believed that the process is asymmetric: A good reputation is hard to win and, once gained, may easily be lost.[2] If this is true for companies, of course, it is also true for BAPEDAL. Indonesian NGO's and their constituencies might well be skeptical of environmental performance ratings from a government which is committed to rapid industrial development. Such reputational considerations have prompted BAPEDAL to adopt a very conservative rating system for PROPER. A polluter must be judged adequate in every environmental dimension to receive an adequate rating. Good or excellent performance in several dimensions is not allowed to compensate for inadequate performance in even one

2.2 Incentive Regulation

Traditional regulation has been plagued by a classic principal-agent problem: Regulators need good data about firms' performance, but firms have clear incentives to withhold such information. In several OECD economies, governments have responded by developing incentive regulation systems for the energy sector.[3] These systems follow traditional practice by penalizing

---

[1] For evidence from North America, see Laplante and Lanoie (1994), Hettige, et. al. (1995a) and Arora and Cason (1994). Recent evidence from developing countries in Asia can be found in Pargal and Wheeler (1995), Hettige, et. al. (1995b), Huq and Wheeler (1992), and Huq, Hartman and Wheeler (1995).

[2] There is little new under the sun in this context. As Mark Anthony noted: "The evil that men do lives after them; The good is oft interred with their bones." (Shakespeare's Julius Caesar, act 3, scene 2)

[3] See Hartman and Wheeler (1994) for a review of research and practice in this area.

non-compliance with regulatory standards. However, they also address the agency problem by rewarding superior performance. This improves the regulator's information by encouraging good performers to identify themselves. It also provides competitive incentives for superior performers to help the regulator identify poor performers, since the latter will be penalized by disclosure.

Conventional incentive regulation has used financial incentives, but reputational incentives address the same agency problems. Since programs like PROPER are new, research will be needed to assess their cost-effectiveness as experience accumulates. A priori, the following assertions about costs and impacts appear reasonable:

Costs: For the public sector, PROPER will cost much less than conventional enforcement because it doesn't rely on time-consuming legal procedures. For some private firms, pursuing Green or Gold status may be very costly. Since the pursuit is voluntary, however, it is reasonable to assume that firms won't undertake it unless the expected gains warrant the costs.

Impacts: Reputational incentives (RI) will generate a different pattern of reponses than either market-based (MB) or command-and-control (CAC) regulation, and under some assumptions RI could generate the greatest overall abatement. Under CAC, polluters in the same regulatory class are all required to meet the same standard regardless of cost. The result is generally convergence to the standard, which may not yield the desired ambient result, and great divergence in marginal cost of abatement across plants. Under MB, polluters will tend toward abatement at equivalent marginal cost, but there will be great divergence in abatement practice. In a pure RI regime, polluters will abate to the point where the marginal cost of abatement is equal to the expected marginal gain in reputation value. Where reputation has no value, polluters may choose not to abate at all. **However, polluters in sectors, communities or markets where reputation has very high value may choose to abate more under RI than under either CAC or MB.** If these are large facilities in pollution-intensive sectors, the result could be overall performance which is also better under RI. At present, we know very little about the determinants of reputational value in developing countries.

2.3 Rating

In numerical or alphabetic form, categorical ratings (grades) are omnipresent in public and private evaluation systems. Some grading systems are dichotomous (e.g., pass/fail); others have many categories. In all cases, however, the number of categories is small by comparison with continuous numerical ratings. Why are limited grading categories so common? Two important considerations are worth noting. First, given its intended use, the grading system should be simple and the implications of any specific grade easily understood. Indeed, grading in a few dimensions serves buyers, sellers and regulatory agencies which need easily-digestible information about relative quality, and are willing to trade speed for precision in making decisions: A few commonly-understood categories are easy to process. In this regard, numerical information in many dimensions is generally suboptimal because the incremental precision doesn't warrant the time and other resources needed to understand the implications. Of course, such a system depends on the credibility of the grading agent.

Second, judgment may be equally important. Grades are generally more than summaries of numerical averages calculated across performance categories. Critical 'add factors' are provided by credible grading agents, who are in a position to judge levels of effort and quality which may not be revealed by simplistic numerical ratings. Those who use grades for decision-making are well aware that expert judgment may be critically important.

Seen in the above light, PROPER's use of five performance ratings seems reasonable. The grading agent in this case is BAPEDAL. Its clients are interested consumers, businesses, stockholders and communities which need simple, credible environmental performance ratings. As we will explain, both measurement and judgment play important roles in grading environmental performance.

2.4 Uncertainty

Even numerical rating is often uncertain, because it is based on stochastic sampling from underlying distributions whose true parameters cannot be known with certainty. Students have bad days on exams; producers occasionally let defective products slip through even if their quality control systems are excellent. When continuous numerical grades in many dimensions are collapsed into simple categories, this problem is compounded. The actual process by which grades are assigned is therefore a function of uncertainty in several dimensions: The sampling variance in numerical scores by category; appropriate weights for combining scores across categories; and the implications of incorrect assignment to very desirable or undesirable categories (which may respectively generate very large premia or losses for some actors).

BAPEDAL faces all these problems in implementing PROPER. Grading polluters' environmental performance requires observation of a complex system comprising many interacting variables. Most important, pollution indicators are measures of central tendency in stochastic effluent samples which are difficult to obtain and subject to large measurement errors. Under existing resource limitations, other important indicators must be derived from indirect observation (e.g., existence of an effectively-operating flow meter is a prerequisite for credible measurement of water pollution). Finally, there is uncertainty in judging the potential impact of pollution loads from individual polluters on neighboring communities and ecosystems.

Taking these factors into account, the PROPER methodology reflects very conservative design and assessment approaches which have been developed to minimize the risk of large grading errors in a system which is manageable and implementable. Since the program is only midway through Phase I, its grading system may well evolve as experience accumulates.

2.5 Summary: Implications for PROPER

To summarize the preceding sections, PROPER is a reputational incentive regulation system based on categorical ratings of polluters' environmental performance. Measurement and expert judgment are used in performance grading, which reflects both a very conservative view of reputation-formation and a risk-minimizing approach to categorical assignment under uncertainty. Adequate performance in all dimensions is necessary for an adequate rating, and extraordinary ratings (good, excellent, very poor) are assigned only after very careful scrutiny. PROPER will

generate substantial abatement at low cost to BAPEDAL if its performance ratings are credible, and if environmental reputation is highly valued by a significant number of polluters.

The following sections of the paper present the methodology which was accepted by the Indonesian Government as the basis for launching PROPER in June, 1995. Section 3 describes Phase I of PROPER: the regulations which form the basis for rating the environmental performance of polluters; the 5-color rating system; and its relationship with compliance requirements. In Section 4, we explain the methodology used to assess the compliance status of polluters. Important topics include: (1) the overall framework of compliance analysis; (2) information requirements; (3) technical guidelines; and (4) the statistical approaches that will be essential for future assessment of compliance status. Section 5 focuses on the critical assignment issues: How is a polluter categorized as Red or Black if found to be non-compliant, and Blue, Green or Gold if found to be in compliance with regulations? Finally, Section 6 provides brief concluding remarks.

3. PARAMETERS OF PROPER - PHASE I

PROPER is expected to develop into a multi-media environmental management program in several phases, most likely following the sequence in which environmental regulations have been introduced for water, hazardous waste and air. Although PROPER is a very innovative program by international standards, BAPEDAL's development strategy is conservative. It has begun with a relatively small, well-articulated program, to be followed by steady expansion to broad coverage in a series of manageable steps with careful interim review and adjustment.

Prior to development of the performance rating methodology, it was necessary for BAPEDAL to address two key issues for Phase I:

- Scope and coverage of the program
- The basis for assessing compliance

3.1 Scope and Coverage: Which Type of Pollution and Sources Should be Included?

Existing environmental regulations in Indonesia cover hazardous wastes as well as air and water pollution. Their compliance requirements vary by type of polluter, generally classified as industrial or non-industrial, stationary or mobile, and point or non-point source. Regulation of hazardous waste and air pollution is very recent, with a Presidential Decree issued in 1994 for hazardous waste and a 1995 Ministerial Decree specifying air emissions standards for stationary sources. Regulation of water pollution has a significantly longer record of development and implementation experience. A 1991 Ministerial Decree (KEP/MEN/03/1991) specifies discharge standards, based on pollution loads for fourteen industries. For the remaining industries, KEP/MEN/03/1991 specifies pollution concentration standards which vary according to water quality objectives in the receiving rivers. The relevant quality categories (A, B, C or D) are identified in guidelines established under the 1990 Presidential Decree on the control of water pollution in Indonesia.

Given its relative depth of experience with regulation of water pollution, BAPEDAL decided to focus on compliance with water regulations in Phase I of PROPER. While it had very limited information on air pollution or hazardous waste, the agency had considerable information on industrial water pollution from two sources: its program for clean rivers (PROKASIH), which was introduced in 1989, and its regulatory enforcement activity (JAGATIRTA) under KEP/MEN/03/1991. Combined with self-monitoring reports from polluters and further inspections of polluters, these information sources were judged sufficient for a careful compliance assessment in Phase I.

3.2 Compliance Assessment: Which Water Pollution Regulations?

Indonesian factories are subject to both national and provincial water pollution regulations. In some cases, the provincial regulations differ significantly from their national counterparts. However, to build public understanding and credibility for PROPER as a national program, BAPEDAL decided to base its Phase I assessment only on national water pollution regulations. Once PROPER is solidly established, it is possible that the methodology will be adapted to incorporate provincial regulations.

3.3 The 5-Color Performance Categories

PROPER is a reputational incentive system with two objectives: To encourage general compliance with the regulations, and to create incentives for pollution reduction in excess of regulatory requirements through adoption of additional end-of-pipe treatment, clean technology, and methods for waste minimization. With these policy objectives in mind, the 5-color rating system has adopted the general criteria shown in Table 2.

While these standards are clear intuitively, objective assessment requires the development and consistent application of definitions for levels of ‘effort’, ‘good housekeeping’, etc. These issues are discussed in Section 5 of the paper.

Table 2

Compliance Status	Color Rating	Performance Criteria
Not in Compliance	**Black**	Polluter makes no effort to control pollution, or causes serious environmental damage.
	Red	Polluter makes some effort to control pollution, but not sufficiently to achieve compliance
In Compliance	**Blue**	Polluter only applies effort sufficient to meet the standard
	Green	Pollution level is lower than the discharge standards by at least 50%. Polluter also ensures proper disposal of sludge; good housekeeping; accurate pollution records; and reasonable maintenance of the waste water treatment system.
	Gold	All requirements of **Green**, plus similar levels of pollution control for air and hazardous waste. Polluter reaches high international standards by making extensive use of clean technology, waste minimization pollution prevention, recycling, etc.

4. COMPLIANCE ANALYSIS

From the criteria in Table 2, it is clear that the first step in assessment must be establishing the compliance status of a polluter. In this section, we address six key issues related to the compliance analysis in PROPER:

- Definition of compliance
- Category assignment process
- Information needed for reliable assessment
- Sources of appropriate information and associated data quality issues
- Use of existing data to assess compliance
- Minimizing errors in category assignment

4.1 Defining Compliance With Water Pollution Regulations

For the purposes of PROPER, compliance with industrial waste water control regulations is defined by two regulations: Presidential Decree PP/20/1990, concerning the control of water pollution; and Ministerial Decree KEP/MEN/03/1991 on the control of industrial waste water. Specific requirements are described in Table 3.

Table 3

Compliance Requirements	Legal Reference
1. Sampling and effluent analyses at least once a month	Article 3.1, KEP/MEN/03/1991
2. Installation of flow meter	Article 5.1, KEP/MEN/03/1991
3. Measuring flow rate daily	Article 5.1, KEP/MEN/03/1991
4. Reporting flow rate data to the agency responsible for monitoring environmental quality	Article 5.2, KEP/MEN/03/1991
5. Reporting true values of pollution	Article 32.1.b, PP 20/1990
6. For industries referred to in KEP/MEN/03/1991, effluent discharged into water cannot exceed the given effluent standards.	Article 7.2.3.a, KEP/MEN/03/1991

Failure to meet any of these six requirements could be judged a legal violation. However, for practical assessment only requirements 1, 5 and 6 can be applied without exception. The rules on monthly reporting of true pollution values are clear. In PROPER, the interpretation of requirement 6 is quite strict: To be judged compliant, a polluter must meet the discharge standard for every pollutant it is expected to control. Even if the standard is violated for only one of many pollutants, the polluter is judged non-compliant. Technically, this judgment is based on sample analysis from six months of pollution data. Non-compliance is defined as violation of the standard for one month or more during the six-month period. This conservative standard was judged necessary on reputational grounds, as discussed in Section 2.

Some ambiguity is introduced into the other three requirements (2,3 and 4) by the provisions for use of a flow meter. Technically, there are two reasons why it is difficult to declare a polluter non-compliant solely on the basis of violations related to flow meter operation in Articles 5.1 and 5.2 of KEP/MEN/03/1991.

1. In Article 5.1 on flow meter installation, the regulations do not define what constitutes an effluent flow meter. It is known that the flow rate can be measured manually with the help of a flotation device and a stop watch. Any polluter could have this equipment, so it would be extremely difficult to establish legal violations.

2. Article 5.1 states unambiguously that all polluters must take flow rate measurements daily. However, practical considerations introduce some ambiguity. Suppose a polluter complies

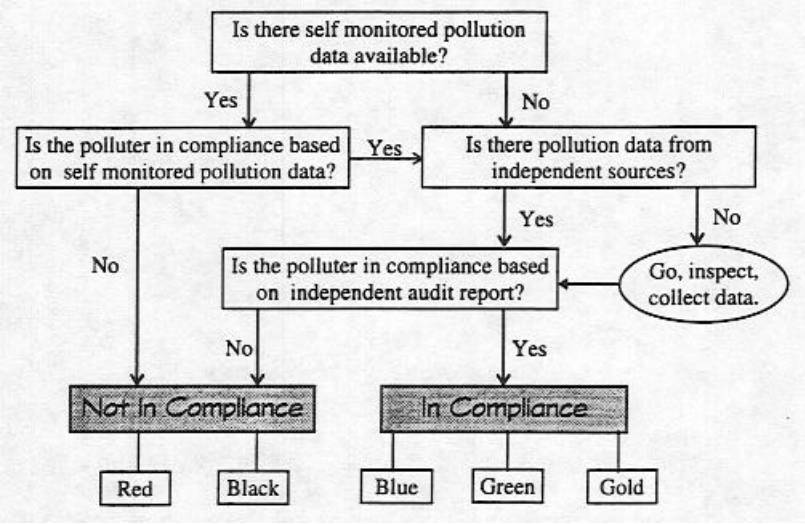
with all legal requirements in Table 3 except for recording 29 flow measurements in a 30-day month? Would it be sensible to judge the facility non-compliant?

By implication, the flow meter provisions should be accorded less emphasis than actual pollution estimates based on appropriate technical and statistical methods. These will be discussed in sections 4.6 and 4.7.

4.2 Steps in Compliance Assessment

Factories are judged non-compliant if they deliberately underreport their pollution, but their actual behavior will be influenced by the probability of discovery. For this reason, the compliance analysis in PROPER relies heavily on pollution data from independent inspections by regulators.[4] As shown in Figure 1, the compliance status of a polluter is first assessed on the basis of self-reported data. If these show a violation of the discharge standards, the polluter is judged non-compliant. If they show no violation, independent inspection and monitoring reports are consulted for verification. If none are available, the plant is inspected by BAPEDAL.

Figure 1



4.3 Information Needed for Compliance Analysis

Specific variables of interest in this context will provide the answers to three key questions:

1. Is the pollution load less than the effluent discharge standards specified in KEP/MEN/03/1991?

[4] However, BAPEDAL recognizes that independent measurements can also be subject to significant error. See Section 4.7

2. Does the polluter meet the requirements of initial compliance with the legislation:

    - Installation of flow meter
    - Sampling and effluent analysis once a month
    - Daily measurement of flow rate.

3. How reliable are the data used for answering Questions (1) and (2)?

Table 4 summarizes the requirements for pollution and production data, which are quite demanding. Much effort in Phase I has focused on database construction, since good information is clearly essential for successful and sustainable implementation of PROPER.

4.4 Sources of Information

BAPEDAL has been collecting pollution data through its PROKASIH program since 1989. After the introduction of KEP/MEN/03/1991 in 1991, its program of legal enforcement (JAGATIRTA) became an additional source of pollution information. When BAPEDAL initiated work on PROPER in 1994, the associated factory inspection work provided a third source of pollution information. Finally, the national regulations require polluters to self-monitor and report their pollution to BAPEDAL on a monthly basis.

Among these data sources, information collected by PROPER teams and JAGATIRTA are considered more reliable than the PROKASIH data. The latter are collected by provincial teams whose competence is sometimes difficult to judge. Thus, the reliability of the PROKASIH data is best assessed on a case-by-case basis. The principal features of the data from differrent sources are summarized in Table 5. The key to assessment in PROPER has been the establishment of a database system which allows for simultaneous comparison of results from all existing sources.

Table 4

Key Questions	Information Needs	Nature of Information
1. Does the plant meet the effluent standards as specified in Appendix I to IV of the regulation KEP/MEN/03/1991 ?	1. What is pollution per unit output of the plant? 2. What is the standard applicable to the plant?	1. Average monthly pollution concentration 2. Average monthly flow rate 3. Monthly output in units specified in KEP-03/MENKLH/II/1991 4. KEP-03/MENKLH/II/1991 standard
2. Does the plant comply with the requirements of Articles 3.1, 5.1, and 5.2 of the regulation KEP/MEN/03/1991?	3. Is there a flow meter? 4. Is flow rate measured and recorded daily? 5. Is the effluent sampled and analyzed once a month?	5. Type of flow meter 6. Number of observations on flow rate per month 7. Number of observations of parameter concentration for the month
3. How reliable are the data used for answering questions (1) and (2)?	6. Are pollution concentration data reliable?	8. Is there an operational waste water treatment system in the plant? 9. What is the sampling method? 10. How frequently are effluents sampled and analyzed? 11. Is the production process batch or continuous? 12. Are data reported for all outlets in the plant?
	7. Are flow rate data reliable?	13. Is the flow continuous ? 14. Is the flow meter reliable? 15. Is the flow meter well maintained? 16. Is the flow measurement taken daily and recorded? 17. Is the production process batch or continuous?
	8. Are production data reliable?	18. Are the units of production consistent with KEP-03/MENKLH/II/1991? 19. Are the production data corrected for intermediate products and by-products? 20. Are production data consistent with the reported capacity of the plant?

Table 5

QUALITY ASPECTS	SOURCES OF DATA			
	Self-Reported	**PROKASIH**	**PROPER**	**JAGATIRTA**
Data Generation Aspects	• Sampling and analysis done by the polluter. Either in-house or an external laboratory or both could be used.	• Sampling and analysis done by the provincial PROKASIH teams. An external laboratory pre-selected by the government officials is generally used.	• Sampling is done by BAPEDAL staff and analysis by a private laboratory.	• Sampling is done by BAPEDAL staff and analysis by a private laboratory.
Positive Features	• Sampled and analyzed at least monthly, so likely to be representative of true pollution characteristics. • If reported properly, data could be legally binding.	• Sampled and analyzed at least once every quarter • Possible to analyze pollution history of plants because information could be available from 1990. Pollution trend provides useful insights into pollution characteristics.	• Most reliable information because sampling and analysis are managed by BAPEDAL staff. • Could be unannounced inspection, so manipulation by polluters would be minimized.	• Most reliable information because sampling and analysis are managed by BAPEDAL staff. • Pollution information is expected to be used for legal sanction, so very detailed and comprehensive.
Negative Features	• Risk of under-reporting pollution is high. • No independent quality control • Sometimes incomplete data are reported. • Sometimes self-reported data are same as the data analyzed by PROKASIH teams	• Technical skills of PROKASIH team inspectors are known to be limited, so sampling errors could be high • Some data reported by PROKASIH are suspicious because of repetition in successive months and quarters. • Sometimes PROKASIH data are the same as the self-reported data.	• Generally representative of short term pollution characteristics.	• Information available only for a limited number of polluters against whom complaints have been registered with BAPEDAL.

4.5 Measuring Compliance with Discharge Standards

Ministerial Decree KEP/MEN/03/1991 specifies discharge standards for fourteen industries. It also specifies the pollutants (by industry type) that polluters are expected to control. The discharge standards are based on pollution loads, specifying allowable discharges of pollutants per unit output of product. More specifically, the structure of discharge standards is:

$$\frac{\text{Quantity of Pollution by Weight}}{\text{Quantity of Industrial Output in Physical Units}}$$

Compliance assessment in this context requires estimates of pollution loads and physical outputs. Concentration and flow rate data are required for pollution load estimation. Since polluters report production data on a monthly basis, compliance with discharge standards is assessed on the basis of average monthly concentration and flow rates. Specifically, for every polluter the quantity of pollution per unit output in a given month (P_m) is calculated as follows:

$$P_m = \frac{\overline{B}_m \times \overline{F}_m \times N_m}{Q_m}$$

where $\overline{B}_m$, $\overline{F}_m$ and N_m are average daily concentration of pollutant, average daily flow rate and number of days of operation of the plant for the month '*m*' respectively. Q_m is the total production in the month '*m*'. If the discharge standard is '*S*', then compliance is assessed as follows:

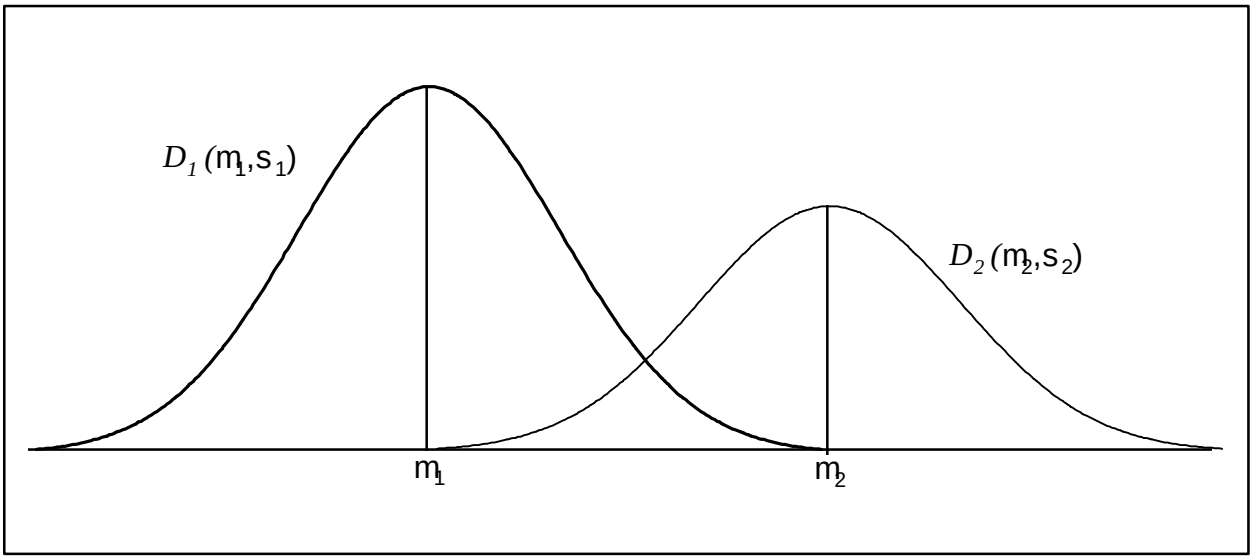
$$\frac{P_m}{S} \begin{cases} \leq 1 \text{ implies compliance} \\ > 1 \text{ implies non - compliance} \end{cases}$$

For PROPER compliance assessment, the value $\frac{P_m}{S}$ is calculated for every pollutant which the polluter is expected to control.

4.6 Reliability of Compliance Assessments

Verification procedures are well-developed in PROPER, but uncertainty remains a challenging problem because water pollution parameters are always judged from sampling estimates. Sampling variation can be substantial, even when samples are drawn repeatedly by competent technicians under optimum conditions. Thus, self-monitoring estimates of pollution levels lower than independent estimates are not necessarily proof of misreporting by polluters. By implication, effluent quality is better represented by a sampling distribution with some mean and variance, rather than by a single scalar. Figure 2 illustrates two hypothetical sampling distributions for a polluter from two reporting sources, with $D_1(\mu_1,\sigma_1)$ self-reported by the polluter and $D_2(\mu_2,\sigma_2)$ generated by independent inspections. Although μ_2 is considerably higher than μ_1 in this case, a judgment of underreporting (non-compliance) should depend on an appropriate statistical test of the hypothesis [H_0: $\mu_1 = \mu_2$]. Sampling variation from both sources would determine the level of confidence associated with the result.

Figure 2



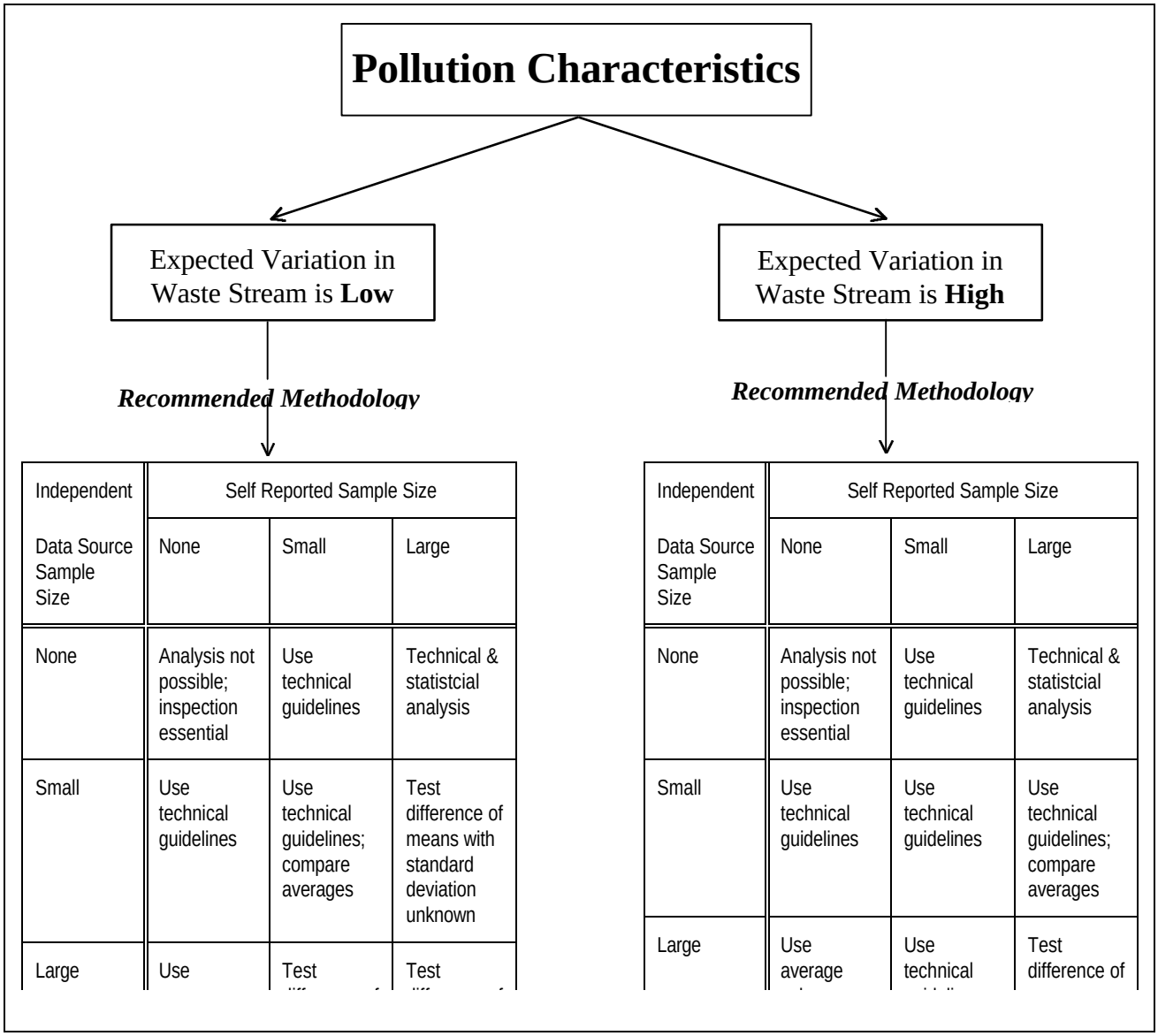
In the case of industrial pollution sources, known sets of technical factors and random effects have significant impacts on sampling results. A sound methodology for testing the reliability of compliance results must therefore incorporate engineering knowledge as well as statistical principles. With this in mind, PROPER's compliance analysis methodology includes:

- Correlation analysis of pollution levels from different sources of data;
- Trend analysis of pollution rather than reliance on spot samples;
- Analysis of the relationship between effluent sampling estimates and the known characteristics of the waste water treatment systems in place;
- Analysis of the relationship between effluent sampling variation and the batch or continuous nature of the production process.

4.7 Summary of the Methodologies for Testing Reliability

The choice of appropriate methodology depends on the sources of pollution and certain properties of the associated sampling data. Figure 3 provides a schematic layout of the methods which are recommended under different conditions. Clearly, a mix of technical and statistical approaches is needed.

Figure 3



Independent Data Source Sample Size	Self Reported Sample Size		
	None	Small	Large
None	Analysis not possible; inspection essential	Use technical guidelines	Technical & statistcial analysis
Small	Use technical guidelines	Use technical guidelines; compare averages	Test difference of means with standard deviation unknown
Large	Use	Test	Test

Independent Data Source Sample Size	Self Reported Sample Size		
	None	Small	Large
None	Analysis not possible; inspection essential	Use technical guidelines	Technical & statistcial analysis
Small	Use technical guidelines	Use technical guidelines	Use technical guidelines; compare averages
Large	Use average	Use technical	Test difference of

As shown in Figure 3, technical guidelines are appropriate when pollution information is based on a limited number of sampling results. As pollution databases build up over time, statistical approaches can be phased in as complements to the technical guidelines.

5. ASSIGNING COLOR RATINGS TO POLLUTERS

The color rating system summarizes the overall environmental performance of a polluter, including treatment of air, water and hazardous wastes. All compliant polluters (i.e., those which meet the requirements summarized in Table 3) qualify as Blue; further evaluation identifies Green and Gold plants. All non-compliant polluters get at least Red ratings; further evaluation identifies Black plants.

5.1 Beyond Compliance: From Blue to Green

As previously noted, it is difficult in some cases to certify that a polluter is fully compliant. So the first requirement for Green or Gold is that all requirements specified in Table 3 be unambiguously satisfied. The second requirement is that all pollution loads be at least 50% lower than the relevant discharge standards in the most recent six months of pollution reports. Polluters meeting these requirements are then scrutinized on the following factors:

- Management and disposal of sludge
- Housekeeping
- Maintenance of pollution records
- Maintenance of the waste water treatment system

To qualify for Green a polluter must demonstrate excellent performance in all four areas. Key elements of sludge management, housekeeping and waste water treatment are evaluated by BAPEDAL inspectors and recorded in photographs for storage in the PROPER database. The quality of pollution record maintenance is judged from the monthly reports submitted to BAPEDAL.

5.2 From Green to Gold

Qualification for Gold status requires satisfaction of all Green requirements, plus a special audit to judge whether a plant demonstrates excellent performance in adoption of clean process technologies, recyclable products, environmentally friendly inputs, and recycling/reuse of materials. These are difficult factors to quantify, and judgment by recognized experts provides the relevant standard. Suffice it to say that this standard is extremely tough: In the Phase I sample of 187 plants, none qualified for Gold.

5.3 Incorporation of Air and Hazardous Waste Assessments

As previously noted, Phase I of PROPER focuses on reliable assessment of compliance status in water pollution because plant-level data on air pollution and hazardous waste are still scarce. Appropriate air and hazardous waste data may be available within a year, but the Indonesian Government decided that it was preferable to avoid long delays in launching PROPER. Until assessments can be based on more complete air and hazardous waste data, the following decision rules have been established for incorporating some assessment of air and hazardous waste performance.

1. If a polluter is assessed as Blue on water pollution, this is the final rating.
2. For Green status, a hazardous waste producer must be rated Green on water pollution and have the operating permit required by the hazardous waste regulation. If this condition is violated, the rating is changed to Blue.

3. For any plant to qualify for Gold status, special audit teams from BAPEDAL would have to evaluate all aspects of its environmental performance and determine the final rating.

In any grading system, critical scrutiny is greater for assignment of extremely good or bad ratings because the stakes are higher. PROPER is normal in this regard, but the current scarcity of information about air and hazardous waste generation raises a potential equity problem: Polluters which are likely candidates for Green status are more closely scrutinized by BAPEDAL than polluters which are clearly no better than Blue. It is possible that such inspections would uncover violations of hazardous waste and air pollution regulations, generating an automatic Red rating for the facility. To combat this 'sampling inequity', BAPEDAL assigns a Blue rating in such cases, with the proviso that it will automatically drop to Red if the plant doesn't reach compliance in six months.

5.4 Non-Compliance: From Red to Black

As noted in Table 2, the distinction between Red and Black ratings depends on assessment of two factors: pollution control 'effort', and degree of pollution damage.

Pollution Control Effort

Effort is judged from easily-observable investments in end-of-pipe treatment or production process alteration. In the context of water pollution non-compliance, Red assignments are maintained for plants which satisfy either of two conditions:

- At least partial installation of primary treatment systems;
- Credible demonstration of pollution-reducing process changes. An example would be a textile plant which has no treatment system but has substituted water-based ink for chemical solvent-based ink in its printing operation.

Plants which satisfy neither condition are given Black ratings.

Pollution Damage

Local damage assessment is at best inexact, even in OECD settings. In the case of PROPER, BAPEDAL is far from having the resources, staff and techniques to support plant-specific damage assessment in all cases. At the same time, it has an interest in punishing clear cases of serious damage with Black ratings. It has therefore settled on a two-step procedure:

1. Plants are considered for Black ratings if they are the object of complaints to BAPEDAL from neighboring communities.
2. Once a complaint is received, it is treated as a potential enforcement problem under the JAGATIRTA program. BAPEDAL staff follow up with an environmental audit of the plant and an assessment of damage from discharges to air, water and land. Black ratings are assigned to plants which are judged to cause serious damage. While this procedure is not

error-free, in practical situations it is often possible to make reasonable judgments. Easy cases in this context would include large polluters with no treatment systems, or scientific evidence that extremely hazardous toxics like mercury or arsenic were being discharged in large quantities.

5.5 The Final Filter: Collective Judgment in BAPEDAL

Because Gold, Green and Black ratings are extraordinary, they pass through a final filter within BAPEDAL. The proposed ratings are discussed intensively by BAPEDAL staff from all divisions. During these discussions any ambiguities are closely scrutinized, and additional information from staff members may be used to improve the accuracy of the assessments. In some cases, the final ratings have been changed as a result of these proceedings.

6. SUMMARY AND CONCLUSIONS

PROPER is a unique program, launched by a new environmental protection institution with very limited resources and information. Using a carefully-articulated assessment system, it assigns color-coded environmental performance ratings to polluters. In Phase I PROPER is focusing on water pollution, with specific consideration of air and hazardous waste pollution in cases where extraordinary ratings (Gold, Green, and Black) are considered.

In its design, PROPER reflects some of the most recent thinking in environmental economics about appropriate incentives for pollution control. It is an incentive regulation system, which uses both carrots and sticks to improve environmental performance. Color coded ratings (Gold, Green, Blue, Red, Black) are used to penalize noncompliant polluters and reward plants which have good environmental performance. In line with the insights of principal-agent theory, the positive incentives also improve BAPEDAL's information base. Good performers come forward to claim their rewards, effectively joining BAPEDAL's effort to identify non-compliant polluters.

As a grading system, BAPEDAL combines numerical measures of critical pollution parameters with expert judgments in dimensions where relevant information is scarce. It follows extremely conservative verification procedures, both to ensure the public credibility of the program and to minimize the risk of color assignment errors. In Phase I, 187 plants have been rated and the summary ratings publicly disclosed as follows:

Table 6

Rating Level	Number of Plants
Gold	0
Green	5
Blue	61
Red	115
Black	6

Soon after PROPER was officially launched in June 1995, all the 187 plants were privately notified, and given until November 95 to improve their performance.

In Phase I of PROPER, BAPEDAL's sequential approach to disclosure has not been criticized. However, questions about its disclosure policy are bound to be raised in the future. Should the agency fully disclose all information which is used for the color ratings? Would this degree of detail actually be helpful to users of the environmental performance ratings? Should all numerical indicators be released, even if judgment factors are hard to summarize for public consumption? If only partial disclosure is feasible, how will credible public representatives get access to complete information in order to certify that the procedures are unbiased? We recognize that these are serious issues, which BAPEDAL will have to resolve as PROPER moves toward broader coverage.

How effective has PROPER been to date? Although it is anecdotal, the initial evidence suggests strong response to the PROPER ratings. Many Red and Black plants have informed BAPEDAL that they intend to come into compliance by December, and no plant has contested its color rating. As predicted by principal-agent theory, many superior performers which were not included in Phase I have volunteered to participate in the next phase of PROPER. Expansion to a larger set of polluters has already begun, and performance data for all PROPER plants will be monitored to determine the impact of the program. Although it is clearly too early to judge the cost-effectiveness of this innovative program, the initial signs are quite favorable.

8. REFERENCES

Arora, S. and T. N. Cason (1994), "An Experiment in Voluntary Environmental Regulation: Participation in EPA's 33/50 Program," **Resources for the Future, Discussion Paper No. 94-10**, Washington, D.C.

Hartman, R., M. Huq and D. Wheeler (1995), "Why Paper Mills Clean Up: Determinants of Pollution Abatement in Four Asian Countries," World Bank, PRDEI (mimeo)

Hettige, M., S. Pargal, M. Singh and D. Wheeler (1995a), "Formal and Informal Regulation of Industrial Pollution: Comparative Evidence from Indonesia and the U.S.," World Bank, PRDEI (mimeo)

Hettige, M., M. Huq, S. Pargal and D. Wheeler (1995b), "Determinants of Pollution Abatement in Developing Countries: Evidence from South and Southeast Asia," World Bank, PRDEI (mimeo)

Huq, M. and D. Wheeler (1992), "Pollution Reduction Without Formal Regulation: Evidence From Bangladesh," World Bank, Environment Department (mimeo)

Laplante, B. and P. Lanoie (1994), "The Market Response to Environmental Incidents in Canada; A Theoretical and Empirical Analysis," **Southern Economic Journal**, 60, 3, pp. 657-672

Pargal, S. and D. Wheeler (1995), "Informal Regulation of Industrial Pollution in Developing Countries: Evidence From Indonesia," **World Bank, Policy Research Department Working Paper No. 1416**