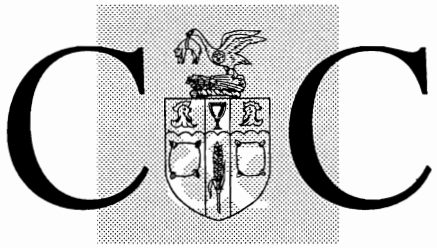
R&D REPORT NO. 27

Authenticity Testing of Rice

April 1996



Campden & Chorleywood
Food Research Association

Campden & Chorleywood
Food Research Association

Director-General
Prof. C. Dennis
BSc, PhD, FIFST

Chipping Campden
Gloucestershire
GL55 6LD UK
Tel: +44 (0)1386 842000
Fax: +44 (0)1386 842100

R&D Report No. 27
Project No. 24357
MAFF Project No. 2A073 (AN1123)

Authenticity Testing of Rice

M.B. Whitworth, M.N. Hall, T. Fearn*,
R.R.L. Seitz**, and N. Franciosi

April 1996

* Dept. of Statistical Science, University College London, Gower Street, London WC1E 6BT
** Now with Department of Physics, University of Aveiro, 3800 Aveiro, Portugal

SUMMARY

The incentive for adulteration of rice by dishonest traders has been recognised by MAFF. Particular authenticity issues concern the adulteration of Basmati rice with other types, and of US long grain rice with rice originating from other countries. As part of previous studies, an image analysis method was identified as suitable for identifying Basmati rice, although it was unsuitable for identifying US long grain rice. The purpose of the current project was to assess the statistical requirements for distinguishing pure from adulterated Basmati samples, using data from the image analysis method, and to develop a full protocol for such analysis of Basmati rice.

The main outcomes of the project were:

- The adoption of new hardware, and development of the necessary new software for image processing measurements, providing higher resolution and improved stability of images.
- Development of calibration procedures to ensure consistent performance of grain measurements, and to facilitate transferability of the method to other instruments.
- Confirmation that the cooking time used for rice is optimal.
- Development of two statistical analysis procedures for the identification of mixed samples, these being:
  (1) A simple to calculate parameter, whose value relative to a cut-off indicates the likely presence or absence of admixture at or above approximately the 20% level.
  (2) A more complex procedure involving determination of the two pure-sample distributions, and their proportions which best fit the measured data. This method also yields a quantitative estimate of the admixture, and of the uncertainty in this estimate.

For the purposes of detecting mixtures, measurement of at least 500 grains is recommended, whereas 50 previously sufficed for the simpler problem of identifying substitution. Analysis (1) has been tested for a limited number of known mixtures at this level. However, most reference samples have only been measured at the 50 grain level, and it is not thought valid to apply analysis (2) to 500-grain samples until the reference samples have been measured more extensively. Analysis (2) has been successfully tested at the 50 grain level, but the uncertainty of determination is high for such small samples. Further development of the statistical method is ongoing, and further testing of reference samples is recommended.

CONTENTS

Page No.

1. INTRODUCTION

Long grain accounts for around 160,000 tonnes out of a total UK rice market of 250,000 tonnes with a retail value of £200 million. Basmati, a class of rice grown in the Punjab region of India and Pakistan, is recognised as a market sector in its own right with about 20% of total rice sales. Basmati can only be grown once a year with a yield half that of other rices; its eating quality cannot be duplicated by growing the same seed in other regions. As a consequence of its scarcity and its popularity in the UK, Basmati commands a high price compared to other rices. The incentive for adulteration was recognised by the Ministry of Agriculture, Fisheries and Food (MAFF) who in 1990 commissioned work at FMBRA (now CCFRA) to develop methods for determining the authenticity of Basmati rice, and by the Grain and Feed Trade Association (GAFTA) whose Rice Standards Section introduced a Code of Practice in January 1992.

The work commissioned by MAFF on the separation of rice proteins by polyacrylamide gel electrophoresis (Scanlon *et al.*, 1991) was only partially successful. For other crops, notably wheat and barley, it is well established practice to identify varieties by biochemical analysis of gene products specific to the seed, particularly the nitrogen storage proteins from the endosperm, known as gliadins or hordeins. These have the merit of being abundant and having many allelic forms that are readily separated into fingerprint patterns by gel electrophoresis or high performance liquid chromatography (HPLC). Unfortunately, rice has adopted a different nitrogen storage mechanism, as have oats, in which prolamine proteins (gliadin type) play a minor role. The abundant glutelin proteins apparently cannot tolerate the many slight mutations evident in the prolamines of most cereals; hence the major rice proteins do not give fingerprint patterns in gel electrophoresis or HPLC. The previous work attempted to derive patterns from the minor prolamine fraction of rice proteins. However, this requires very high resolution and sensitivity which is difficult to reproduce in routine operation. Furthermore, Basmati is defined by origin as well as genotype. Gene-specific components would be expected to be consistent across samples of the same genotype from different origins. As a result of these considerations, further research focused on objective physical methods.

CCFRA has developed three tests (image analysis, rapid visco-analyzer (RVA) and near infrared spectroscopy (NIR)) for evaluating the authenticity of Basmati rice (Osborne *et al.*1993b). Each of the tests is based on a physical measurement which is characteristic of both the genetic make-up of Basmati and of the environment under which it was grown. In that study, 132 samples of milled rice of known provenance were collected, of which 89 were tested by all three techniques to provide a database against which samples of unknown provenance could be compared.

A subsequent study (Whitworth *et al.* 1995) to validate these tests, using a further set of 95 samples, confirmed the success of image analysis for distinguishing Basmati samples from other types of rice in nearly all cases, although this test was not capable of determining the authenticity of US long grain rice.

On the basis of the success of image analysis for distinguishing pure samples of Basmati from other types of rice, it was identified that the technique also has potential for discrimination of pure Basmati samples from those to which other rice types have been added. The purpose of the current project is to establish a protocol for the determination of the quantity of non-Basmati rice present in admixtures to Basmati rice. This requires determination of the potential level of addition that could be detected in this manner, and the sample size required to achieve this with a suitable level of confidence. Additionally, the results of previous studies had revealed a bias between the measurements made in the two studies. It was therefore required that the protocol should include safeguards against future bias, and should include provision for replication of results on other equipment.

Image analysis of cooked rice exploits the characteristic shape change which Basmati rice undergoes during cooking. Unbroken grains from the sample are cooked, either loose in a beaker (Osborne *et al.*, 1993b), or in individual holes in a special tray placed into the beaker (Whitworth *et al.*, 1995). The cooked grains are then placed on a light box and a video camera is used to capture their silhouette images for computer processing. First, the computer makes a correction for background lighting, then it thresholds the image to identify dark regions and stores those measurements consistent with rice grains. For each grain, several size and shape measurements are combined into a single value (Rice Parameter 3, or RP3),

which is greater for non-Basmati than for Basmati grains, and can be used for discriminant purposes.

2. MATERIALS AND METHODS

2.1 Sample collection

Databases of 132 and 95 samples of rice had been previously collected for two projects (Osborne *et al.*, 1993b, Whitworth *et al.*, 1995), and have been previously referred to as databases 1 and 2 respectively. For this project, samples were selected from database 2, which comprises a total of 19 Basmati, 9 Indian non-Basmati, 43 US long grain, and 24 other rice samples obtained from a variety of sources. Each sample of rice had been milled (whitened) prior to receipt at CCFRA. Full details of database 2 are provided in a previous report (Whitworth et al., 1995). Measurements had previously been made of all samples in database 2. These data were used in this study to determine the requirements and statistical analysis procedures for testing of mixtures, and to select typical samples for further study. Although the analysis procedure had been modified in this study, it was expected on the basis of previous experience that any difference in results between procedures could be approximated by a constant bias in Rice Parameter 3. By selection of samples with a wide range of mean RP3 values, this bias correction has been calculated (see section 3.2), and has obviated the need to re-test all of the samples. In addition to use of data previously generated, new measurements were also made on mixed samples created (section 2.6) specifically for the purpose of testing the new protocol. The number of grains examined in the case of these samples was 500 (see section 3.3.3 for the reason for this increased number of grains).

2.2 Measurement procedure - introduction

Rice grains were boiled for analysis on the premise that the difference in size and shape of Basmati and other rices is greater when they are cooked than when uncooked. A measurement of each grain is made, which is related to its degree of elongation, and can be used to assess its type on a statistical basis.

2.3 Cooking procedure

In a previous study (Whitworth *et al.*, 1995), grains were placed in a specially designed tray in order to allow each grain to be individually identified for comparison of NIR and image analysis data. For the present study, and protocol developed within it, only image analysis measurements were required, and the cooking method reverted to that used in a previous study (Osborne *et al.*, 1993b), in which grains are loose in a beaker during cooking, permitting greater numbers of grains to be cooked at one time.

Figure 1 shows the apparatus used to cook rice grains. The procedure is described in more detail in the Appendix. The rice is cooked in a beaker of distilled water, suspended in a boiling water bath. Tests were conducted to establish the level of discrimination of rice types achievable for various durations of cooking (see section 3.1), and a time of 10 minutes has been adopted for the protocol, as used in previous studies.

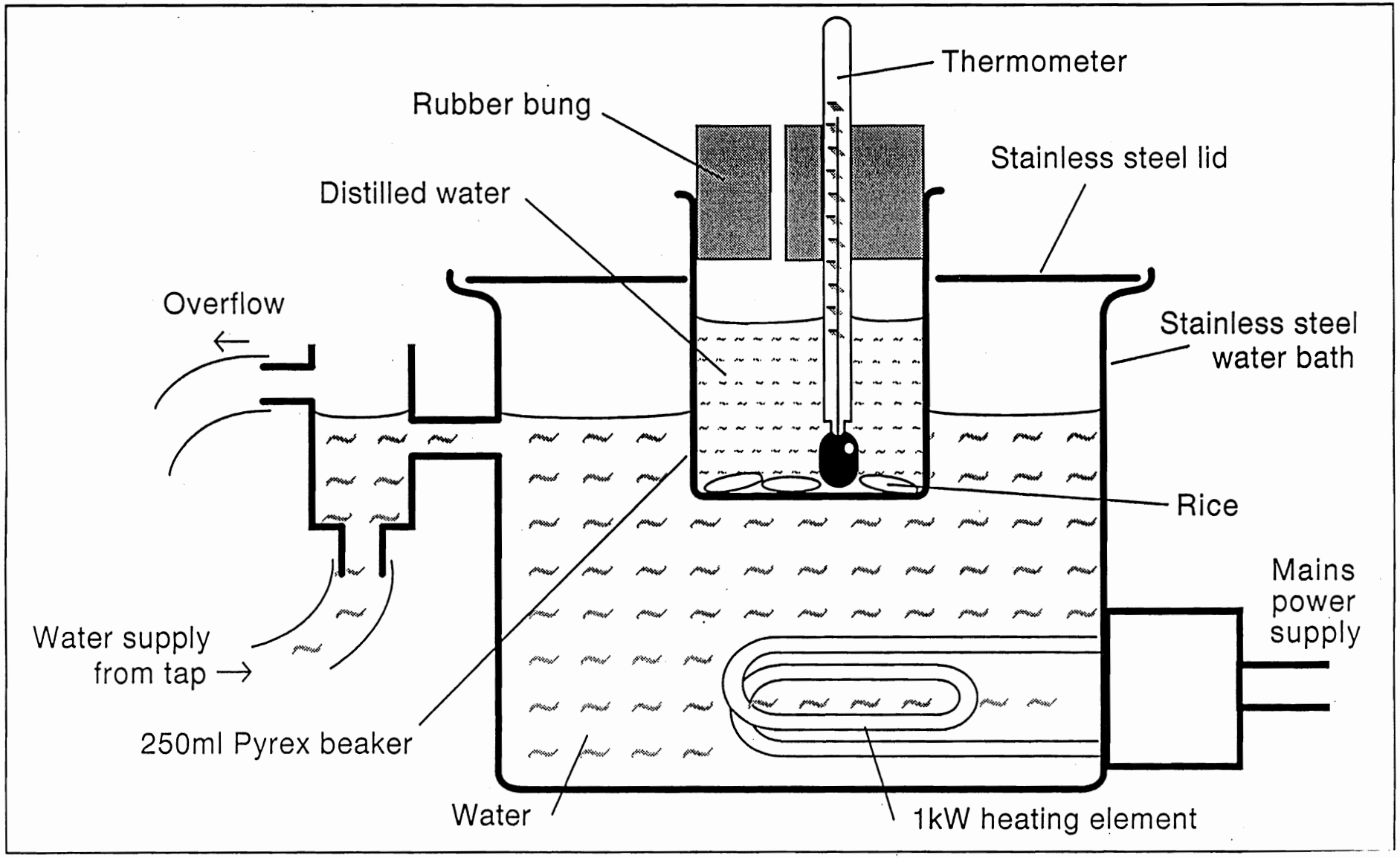
2.3.1 The effect of cooking time on discrimination ability

In a previous study (Scanlon *et al.*, 1991), a cooking time of 8 minutes was adopted, since grains were found to be poorly discriminated on the basis of a length to breadth ratio for cooking times of only 5 minutes, but were found occasionally to rupture when the cooking time was increased to 10 minutes. During previous development of the image analysis test used for the current study, the cooking method was changed to use a boiling water bath, rather than placing the rice in directly heated boiling water, and a cooking time of 10 minutes was adopted. This procedure was chosen on the advice of GAFTA members that it was similar to a procedure used by one of them to cook rice for other testing, and was designed to produce a texture suitable for consumption.

Although this is suitable for complete cooking of the rice, and better discrimination between rice types had been observed for such cooked grains than for uncooked grains, it remained to be established whether the chosen cooking time was optimal for the purposes of discrimination by the measurement (RP3) used for this study.

Figure 1

Apparatus used to cook rice grains



Tests were therefore conducted in which 50 grain subsamples of 3 Basmati rice samples and 3 non-Basmati samples were each cooked for varying amounts of time, the selected times being:

5 minutes, 7.5 minutes, 10 minutes, 15 minutes.

Each sample was analysed by image analysis, and the degree to which the Basmati samples were discriminated from the non-Basmati samples according to the discriminant function RP3 was determined for each of the cooking times. The rice samples chosen were examples of those which had previously provided a wide range of mean RP3 values (for a 10 minute cooking time). They included examples of those previously well discriminated, and those previously poorly discriminated.

Results are given in section 3.1.

2.4 Image analysis

A JVC CCD camera was mounted above a light box on a photographic enlarger stand. The camera generates an RGB and a PAL output, of which the PAL signal was used. The field of view was approximately 30mm x 40mm. In previous studies, a Solitaire computer, manufactured by Seescan Imaging, had been used for image acquisition and analysis. This captured monochrome (the brightness component of the PAL signal) images to a resolution of 512 x 512 pixels x 7 bits. For the present study, a Spectra decoder (Synoptics Ltd, Cambridge) was used to convert the PAL signal to an RGB signal. Images were instead grabbed to a resolution of 768 x 576 pixels x 8 bits, from the green component of the signal, using a Sprynt 40MHz frame grabber also supplied by Synoptics Ltd and installed in a 66MHz, 486 PC computer. Images were analysed using library programs developed and executed within the Semper for Windows application program, also available from Synoptics Ltd.

2.4.1 Calibration

The need for calibration

Previous studies had shown a drift in the value of the function RP3, which is a discriminant function calculated for each grain. RP3 is a function of several grain measurements, and the principal source of the drift was the measured value of the perimeter of each grain, the mean value of which had increased between the first and second studies. The exact cause of this increase is uncertain, but a likely possibility is a slight lack of synchronisation between the two fields of alternate scan lines comprising the PAL image, resulting in a convoluted boundary to each measured object.

Use of reference objects

In an attempt to avoid recurrence of this problem with the Seescan Solitaire computer used for previous studies, the Synoptics computer system described above was used, which experience has suggested is able to generate a more stable image. As a safeguard against recurrence of this problem, or of other causes of bias in the measurements, the procedure was expanded to include measurement of reference objects. These are intended to resemble rice grains, and to be measured in the same manner as these, prior to analysis of rice samples. Any systematic error likely to affect measurement of rice grains is also likely to alter the measurement of the reference objects, and thus to be detectable.

The chosen reference objects were nylon cylinders, formed by machining off the thread from a piece of nylon studding. The cylinders had diameters of 2.96 ± 0.03 mm, and lengths of 9.06 ± 0.02 mm, selected to be similar to the dimensions of Basmati rice grains. Nylon was chosen as a material, since it has a similar transparency to cooked Basmati rice. This was established by placing the cylinders on the lightbox alongside rice grains, and comparing the brightness of their digitised images due to the transmitted light. Of several polymers tested, nylon gave the closest match to rice for cylinders of similar size. It is important that the appearance of a reference object should be stable over time. Although nylon was chosen for the similarity of its transparency to cooked rice, it is known that polymers can degrade with time, particularly in response to light. Therefore, the cylinders are kept in an opaque container when not in use.

For the purposes of creating secondary reference samples, it is intended that the primary reference samples should first be used to test the correct calibration of CCFRA's computer system. This system can then be used to make measurements of the secondary references, which should be used as their standard values, and should be matched to within accepted tolerance when future tests of rice samples are made, possibly on different instruments.

Calibration procedure

The procedure for setting up the image analysis system and operating the rice analysis program is described in detail in the Appendix.

The brightness of the image is adjusted by means of the camera aperture, to a standard brightness, assessed with the aid of false colour contours of brightness displayed on screen.

The magnification of the image is controlled by adjusting the height of the camera above the lightbox, and is measured by placing a ruler on the lightbox, and measuring the spacing in image co-ordinates of two points a known distance apart on the ruler. To minimise the opportunity for variations in the analysis procedure, the magnification is adjusted until it lies within a narrow range, chosen to be approximately equivalent to a variation of ± 1 mm in the camera height.

When the camera has been adjusted and focused, an image of the empty lightbox is grabbed and stored. This is used to normalise all subsequent images acquired for the rice sample, including those of reference objects. The normalisation is intended to provide correction for any variations in image brightness that may occur between samples, despite the initial adjustment of the camera aperture. It is also intended to provide correction for variations in brightness across the field of view.

At least 3 nylon reference cylinders are placed on the lightbox, and measured in the same manner as for rice grains. For each cylinder, the value of RP3 (the discriminant function on the basis of which rice grains are assessed for similarity to Basmati) is compared to a previously measured value for that cylinder (the cylinders are identified by numbers written on their ends).

2.4.2 Measurement of rice grains

For analysis, rice grains are placed on the lightbox with their symmetry planes horizontal, such that they do not abut, and such that they do not overlap the edges of the video frame. Typically 10 grains are analysed at a time. After their image has been grabbed for analysis, they are replaced with further grains, and the process repeated until all grains have been analysed. For more than about 100 grains, the manual placement of them on the light box is the most time consuming element of the test.

Images are analysed by first dividing by the stored background image of the empty lightbox, and scaling to a brightness range of 0 (black) to 128 (white). Grains are then identified by binary thresholding at a grey level of 116. Objects darker than the threshold are identified as candidate rice grains, and are measured to determine their length, breadth, perimeter, and area, which are scaled to units of millimetres according to the previously determined magnification factor. The length is determined as the feret (ie caliper measurement) measured in the direction of the principal moment of area, and the breadth is the feret measured perpendicular to this.

The function, RP3 was also calculated for each grain image. This parameter has been developed in previous studies as a function which provides good discrimination between Basmati (grains of which tend to have a low mean value of RP3) and other rice types (which tend to have higher mean RP3 values). The exact functionality of RP3 is a commercial secret; it is based on the fact that Basmati grains have a high length to breadth ratio, but unlike this ratio, it makes allowance for the possible curvature of grains, which tends to increase their breadth and thus reduce their length to breadth ratio.

All objects of area greater than or equal to 5 mm^2 are considered to be rice grains. Small fragments of grain and pieces of dust do not satisfy this criterion, and their measurements are rejected. Values of length, breadth, area, perimeter, and RP3 for each reference sample and for each rice grain are automatically saved to a computer disc for further analysis.

2.5 Analysis of data

The variation in RP3 measurements between individual grains in the same rice sample and between rice samples was studied by analysis of variance. The results and their implications for the detection of adulteration are reported in section 3.3.

Two approaches to detecting adulteration were studied. Both attempt to detect mixtures of overlapping populations, which is what would typically result from admixture.

The simpler of the two approaches involved calculating measures of within-sample spread designed to be sensitive to the type of mixtures of interest. The measures were calibrated using the database of pure samples and their performance tested on mixtures generated mathematically from the database, and on a small number of actual mixtures. These results are reported in section 3.4.

The more complex but potentially far more powerful approach involved fitting mixtures of Gaussian distributions to the observed data for the suspect sample. The methodology used was Bayesian (Gelman *et al.*, 1995, chapter 16), with prior distributions for the means and standard deviations of the two potential components of the mixture derived from the database of 95 samples. Markov chain Monte Carlo techniques were used to simulate samples from the posterior distributions of the mixture parameters given the observed data. In particular the probability distribution of the proportion of non-Basmati rice in the mixture can be displayed, and any required probabilities (eg the probability of there being more than 20% non-Basmati rice in the sample) can be calculated. Full technical details of this approach will be provided in a research report (Fearn, 1996). The results of applying the method to actual mixtures are reported in section 3.5.

2.6 Creation of mixed samples for assessment of method

In addition to the simulation of mixed samples using single grain data from previously measured pure samples, true mixed samples were also created for the purposes of assessing the performance of the test protocol. Such samples were created at four different levels of

admixture (10%, 20%, 30%, 50% by mass) of a non-Basmati to a Basmati sample. Although samples of 46 grains each had been adequate for the purposes of characterising the distributions of grain results within pure samples, larger samples are required for the purposes of detecting admixture (see section 3.3.3), and mixed samples were therefore tested at the sample size of 500 grains specified in the protocol (see Appendix). Because of the greater sample size, and the several levels of admixture, it was impractical to test the protocol on an extensive number of samples from the database. Instead, mixtures were made from two Basmati and two non-Basmati samples, each of which in isolation yielded a distribution of RP3 values typical respectively of these classes. Each mixture was formed by combining one of the Basmati with one of the non-Basmati samples, making 4 sample combinations in total, each at four levels of addition.

The samples used for the creation of mixtures were:

Basmati			**Non-Basmati**	
MR94/423	India		USA Long grain	Sainsburys
MR94/423	India	Flying Horse	USA Long grain	Tilda

The non Basmati adulterants were purchased commercially in addition to the original database.

The quantities of each sample used in the mixtures were determined by weight. Each constituent sample was representatively subsampled by iterative sample division using a Gilson MiniSplitter SP3, to provide appropriate quantities to yield 80g of each mixed sample. Although mixtures were formulated by weight, the number of grains in each component of the mixture was also counted using an automatic seed counter. The two components of the mixture were then combined, and thoroughly mixed by hand.

The mixed sample size of 80g is greater than that required for analysis. This was therefore subsampled by successive sample division until the sample size had been reduced to the 500 grain size required for testing. This procedure ensured that the actual proportions of the two components in the tested subsample were subject to a sampling error, and thus provided a more realistic assessment of the true performance of the testing protocol.

2.7 Transferability of method

The method as developed and used in this report is described in the Appendix. Full details of the procedure for fitting mixtures of Gaussian distributions to the data will be provided in a research report (Fearn, 1996). Some details of the procedure (e.g. models of some equipment) are unlikely to be crucial to the replication of results generated by this protocol, but it is recommended that the effect of any alterations to the procedure should be carefully assessed. In particular, any alternative version of the method or additional set of apparatus should be tested to ensure that measurements of reference objects (see section 2.4.1) are consistent with those made with the system described here. Procedures for generating secondary reference objects are described in section 2.4.1.

The procedures for measuring length, breadth, area, and perimeter of objects in binary images are well known, and many image processing application programs (including that described here) provide standard routines for such measurements. The parameter RP3 is a function of these values, but the functionality remains a commercial secret. Those wishing to make use of this function are advised to contact CCFRA to negotiate terms for its use, or for CCFRA to make the calculations on their behalf.

The database rice samples used in this study were collected as part of a project funded by MAFF, to whom requests for their use should be addressed. Although the samples are currently stored by CCFRA, it is believed by the authors that the procedures for storage of, and access to such sample sets are currently under review by MAFF.

3. RESULTS

3.1 Effect of cooking time on discrimination ability

50-grain subsamples of six rice varieties were each cooked for four different lengths of time, and measured by the image analysis procedure. Means and standard deviations of RP3 for each 50-grain sample are shown in Table 1.

Table 1: Effect of cooking duration on the distribution of RP3 values

Rice sample (MR94/-), variety	Origin	Cooking time	5 minutes	7.5 minutes	10 minutes	15 minutes
147 **Basmati** PAK	Pakistan	Mean RP3 (s.d.)	25.0 (4.2)	26.7 (4.1)	28.9 (5.2)	34.6 (3.8)
432 Pusa **Basmati**-1	India	Mean RP3 (s.d.)	26.9 (4.2)	30.3 (4.9)	29.4 (5.5)	37.2 (5.5)
434 Dehradum **Basmati**	India	Mean RP3 (s.d.)	31.2 (4.3)	32.0 (4.9)	33.9 (5.9)	37.1 (4.6)
125 Guarav	India	Mean RP3 (s.d.)	33.0 (4.4)	35.9 (5.6)	37.1 (4.5)	42.2 (4.2)
150 Icaro	Italy	Mean RP3 (s.d.)	36.2 (4.9)	38.5 (4.2)	41.3 (4.5)	43.8 (4.4)
1923 Katy	Mississippi, USA	Mean RP3 (s.d.)	34.9 (6.5)	38.1 (5.2)	41.2 (5.9)	40.2 (5.9)
Mean of Basmatis			27.7	29.7	30.7	36.3
Mean of others			34.7	37.5	39.8	42.1
Separation			7.0	7.8	9.1	5.8
Pooled within sample s.d.			2.5	2.1	2.6	1.6
Separation / s.d.			2.8	3.6	3.5	3.5

It can be seen from the separation / s.d. ratios that the level of discrimination is similar for cooking times of 7.5, 10, or 15 minutes, but that a slightly inferior discrimination is achieved for a time of only 5 minutes. Although a cooking time of 7.5 minutes might result in a slightly quicker test than the 10 minutes previously used, and could be expected to result in less breakage of grains during cooking as observed by Scanlon *et al.* (1991) for an 8 minute

cooking time, it provides no discriminatory benefit. A cooking time of 10 minutes was therefore retained, in order that continued benefit could be drawn from the results of previous studies using this procedure. A 10 minute cooking time has therefore been documented in the protocol described in the Appendix.

It can be calculated from the results that the value of RP3 increases with cooking time at a rate of approximately 0.8 units per minute. Deviations of the order of 1 unit in RP3 are known to be significant in the determination of Basmati authenticity, and it is therefore important that the cooking time (including the time for which the rice remains in hot water after removal from the water bath) is controlled to within a precision of 30 s or better. This has been noted in the description of the protocol.

3.2 Determination of bias

It was noted by Whitworth *et al.* (1995) that a bias had occurred in the mean values of RP3 relative to values established by a similar method in a previous study (Osborne *et al.*, 1993b). In the current study, a revised protocol has been adopted in order to detect whether further biases occur in future. However, it was anticipated that the adoption of a new protocol and computer system might have introduced a bias relative to one or both of these previous studies. A selection of samples from database 2 (as used by Whitworth *et al.*, 1995) were therefore analysed in order to determine the value of this bias, and to enable test criteria to be established relative to the larger sample set previously measured.

Ten samples were chosen for re-analysis by the new procedure, and are listed in Table 2. These samples (2 Basmati, 1 Indian non-Basmati, 4 USA, 2 European and 1 South American) were chosen randomly from the 5 categories listed to give a representative sample of the database.

Table 2: Samples re-analysed to assess the effects on RP3 of the revised analysis procedure

Sample number (MR94/-)	Origin and type/variety	Sample number (MR94/-)	Origin and type/variety
81	French Guyana, L7	425	India, **Basmati**
87	Camargue, Dwarf japonica (USA variety)	1916	Mississippi, Kaybonnet
125	India, Guarav	1918	Texas, Gulfmont
130	Italy, Thaibonnet	1928	Texas, Newbonnet
147	Pakistan, **Basmati** PAK	1935	Louisiana, L202

Analysis of the results revealed:

(1) A consistent shift of -1.5 units in RP3 from the old protocol to the new. Adjusting for this in comparing old and new results is a simple matter.

(2) An increase in within-sample standard deviation of RP3 of around 25%. The reason for this is unclear, but as with the shift in mean, it was a consistent effect. Again, an adjustment for this can be made quite easily, although a reduction in variability would be preferable to improve discrimination capability.

3.3 Between and within-sample variation in RP3 and its implications

These results are based on an analysis of between- and within-sample variability in RP3 measurements previously made (Whitworth *et al.*, 1995) of the 95 samples (19 Basmati, 76 non-Basmati) in the database. For most of the samples, 46 grains were measured.

3.3.1 Between-sample variation

The mean RP3 measurements for the 19 Basmati samples were approximately normally distributed with mean 33.5 and standard deviation 1.9 units. The mean RP3 measurement for

the 76 non-Basmati samples appeared to come from a distribution with heavier tails than the normal, although a normal distribution with mean 42.9 and standard deviation 3.0 units is a fair approximation. These two distributions overlap, but most Basmati samples would be expected to have a mean RP3 less than 37 and most non-Basmati samples to have a mean greater than 37. Hence the ability of RP3 to detect most cases of substitution, as concluded by Whitworth *et al.* (1995).

3.3.2 Within-sample variation

The within-sample distributions of RP3 measurements were, with the exception of a few outlying measurements, consistent with being Gaussian. The pooled within-sample standard deviations of the measurements on individual grains were 3.5 for the Basmati and 3.4 for the non-Basmati samples.

3.3.3 Implications

The relatively large within-sample variation in RP3 means that there is considerable overlap in the individual grains between Basmati and non-Basmati samples that are quite clearly separated in their means. For example, a Basmati rice with a mean RP3 of 35 units, easily separable from non-Basmati samples by its mean, would typically have one third of its grains with an RP3 measurement of greater than 37, the Basmati cut-off.

The implication of this is that attempting to detect admixture by setting a cut-off and declaring the sample adulterated if more than a few grains exceed the cut-off is not a promising approach. The cutoff would need to be set so high to allow for the within-sample variation in genuine Basmatis that many non-Basmati grains would fall inside it. Instead it will be necessary to use more of the information in the measurements on the suspect sample, and in particular to try to detect the presence of a mixture distribution.

A further implication of the overlap between the Basmati and non-Basmati populations is that a larger number of grains are required in order to identify mixed sample populations (i.e. adulteration) than to classify pure samples (i.e. substitution). This is because of the

requirement to distinguish real deviations in populations from statistical noise. Additionally, larger numbers of grains should be drawn from mixed samples in order to provide sufficient sampling of the minority components within them. A sample size of 500 grains was therefore adopted as a suitable compromise between these considerations and the practicalities and cost of testing large numbers of grains.

3.4 Simple approach to detecting mixtures

The idea of this approach was to find a simple statistic, to be calculated from the individual grain measurements on a sample, that is sensitive to the presence of a mixture. A number of possibilities were investigated, mainly based on the sample percentiles. If the measurements (n=46 for most of the database samples) are ordered from smallest to largest then the p th percentile Q(p) is the value below which p% of the measurements lie, interpolating if necessary. For example, Q(50) is the median measurement. The most promising statistic of those investigated was a value, T, defined as:

$$T=Q(90)-Q(50),$$

which is a measure of the spread of the upper half of the distribution. Because RP3 values for non-Basmati samples are greater on average than those for Basmatis, it is the upper tail of the RP3 distribution for a mixture that will be distorted by the presence of the adulterant (unless it is present in a sufficient proportion to shift the whole distribution). To enable detection of this latter case as well as moderate adulteration, it is suggested that T be plotted against the median, Q(50). Figures 2-5 show plots of this form for the database of pure samples and a number of simulated mixtures. Each mixture was simulated by randomly selecting a Basmati sample, randomly selecting an adulterant, and then mathematically generating a mixture by combining randomly selected single grain measurements in the specified proportions from their respective datasets.

Figures 2-5

Plots of T (the difference between the 90th percentile, Q(90) and the median Q(50) RP3 values) against Q(50) for all 95 samples in the database of pure samples, and for a number of simulated mixtures. Proportions of admixture used are 10, 20, 30 and 50% in Figures 2 to 5 respectively.

Key:	Triangles:-	Pure Basmati samples
	Circles:-	Pure non-Basmati samples
	Crosses:-	Mixtures

Figure 6

Plot of T against Q(50) as for Figures 2-5, for measurements of real mixed samples.

Key:	Solid triangles:-	Pure Basmati samples
	Solid circles:-	Pure non-Basmati samples
	Crosses:-	10% admixture of non-Basmati
	Open triangles:-	20% admixture
	Squares:-	30% admixture
	Open circles:-	50% admixture

Figure 2

Database pure samples (Basmati: ▲; other: •), and simulated 10% admixtures (x).

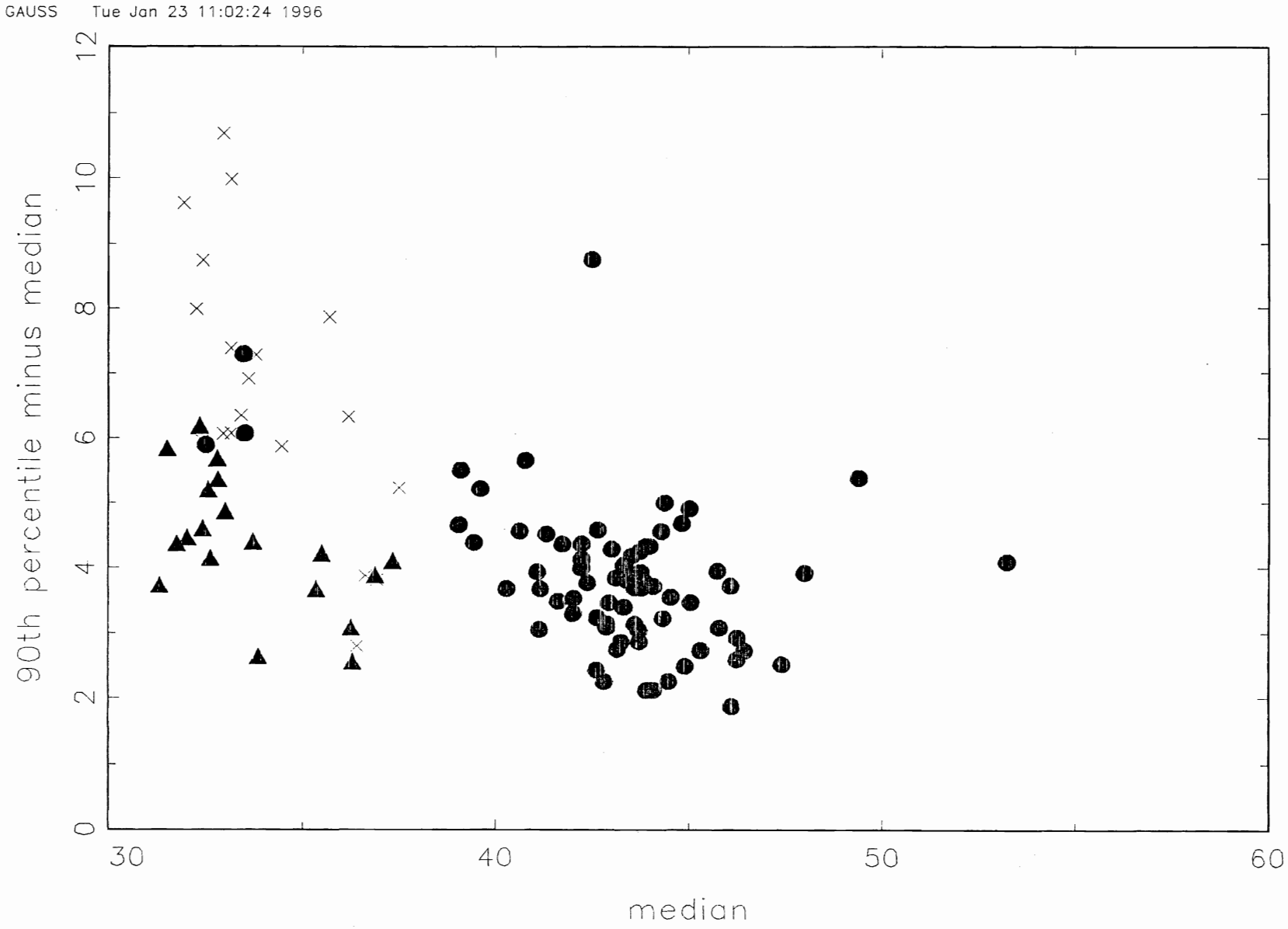


Figure 3

Database pure samples (Basmati:▲; other:•), and simulated 20% admixtures (x).

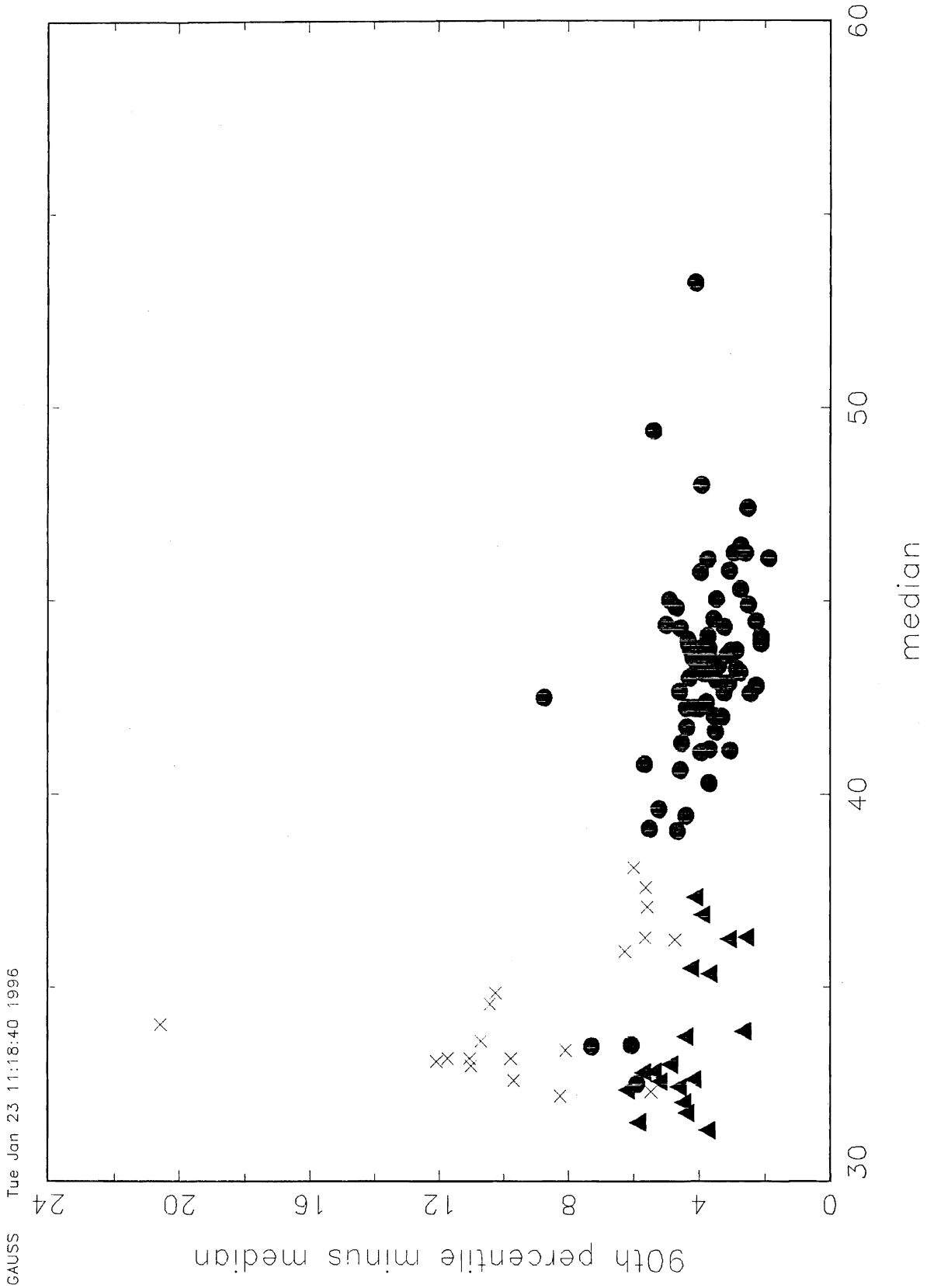


Figure 4

Database pure samples (Basmati:▲; other:•), and simulated 30% admixtures (x).

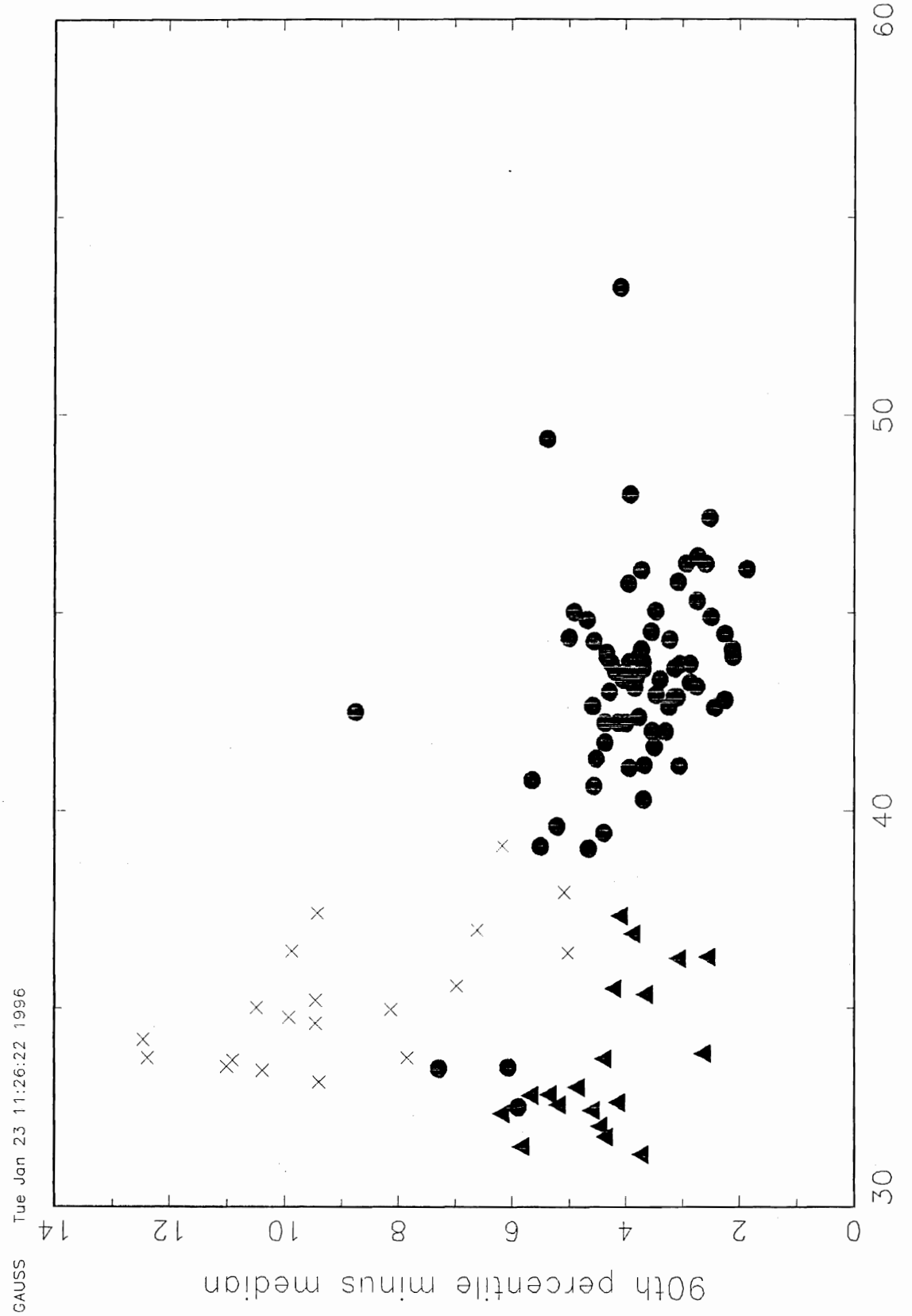


Figure 5

Database pure samples (Basmati:▲; other:•), and simulated 50% admixtures (x).

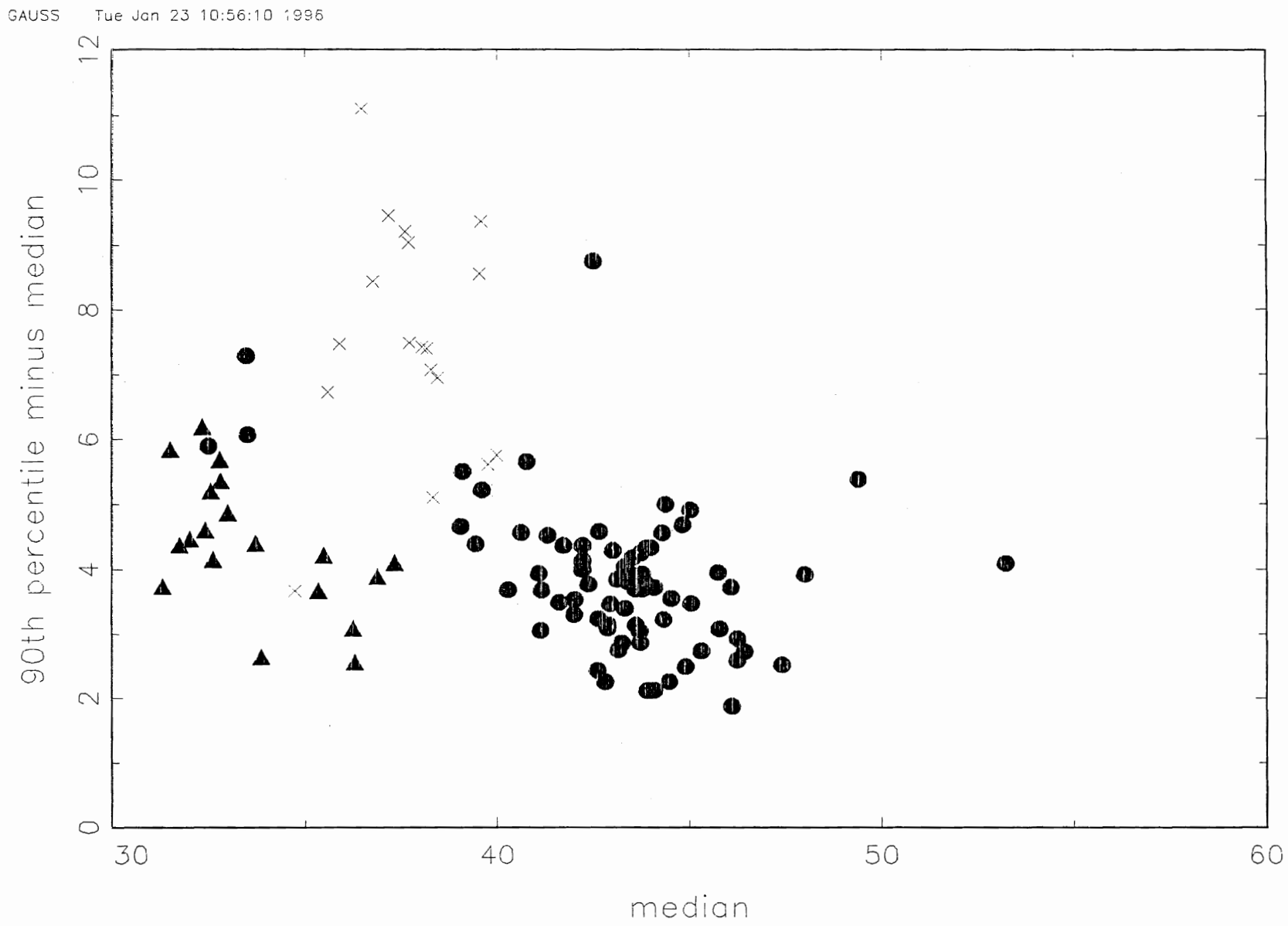
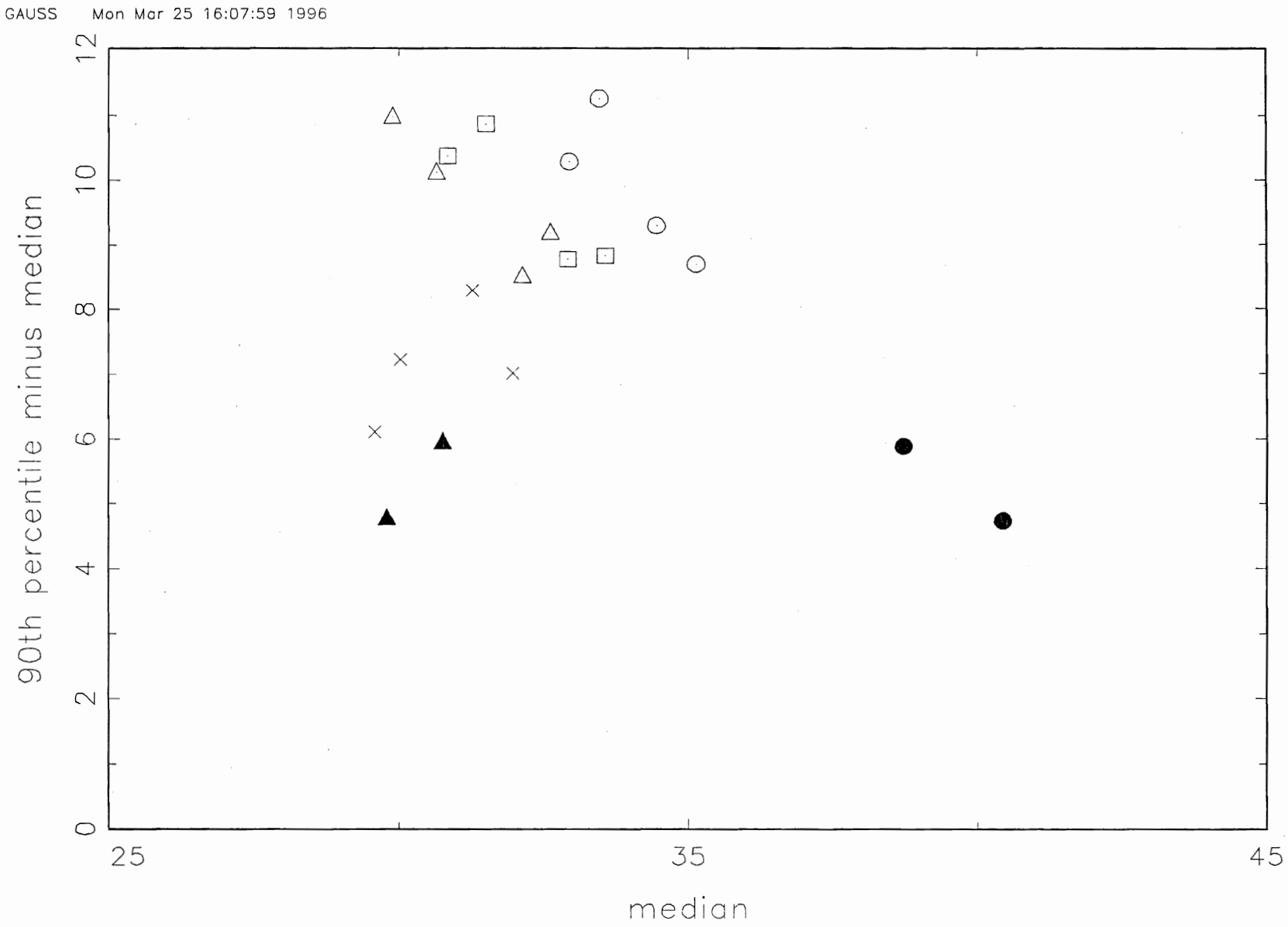


Figure 6

Pure samples (Basmati: ▲; other: •) and real mixtures thereof.



Inspection of the figures leads to the following conclusions. Under the old protocol, the maximum T for any pure Basmati is just over 6 units. About half of the samples with 10% addition, increasing to most of those with 50% addition, are above a cut-off set at about 6.5, and so the statistic does seem capable of detecting admixture. It can also be seen that the medians increase as the proportion of non-Basmati in a sample increases. Some of the heavily adulterated samples are detectable because the whole distribution has been shifted out of the Basmati region. Under the old protocol, the greatest median RP3 value measured for the database Basmati samples was 37.3, and heavy adulteration, or substitution of non-Basmati samples, would therefore be detected if the median RP3 exceeded a cut-off set slightly above this level. The problem with the above approach is that it does not allow the level of admixture to be quantified, and would be hard pressed to distinguish between, for example, 15 and 30% addition. Such distinctions may be crucial for distinguishing between legitimate and fraudulent addition. It should be noted that one of the non-Basmati reference samples (Italy, no 155) itself shows characteristics of being a mixture. It is however one of a small number of samples from which only 23 grain measurements were available.

Figure 6 shows a similar plot for the measurements of real mixed samples, prepared as described in section 2.6, and measured with the new protocol. Samples of 500 grains were measured in this instance, instead of the 46 used in Figures 2-5; however, this should not change the expected value of the statistic, only its precision. Because these samples were measured under the new protocol, for which the standard deviation increased by 25% (see 3.2), the cutoff in T of just over 6 units derived from Figures 2-5 should be increased to 8 units. All but the 10% samples would be detected. To detect heavy adulteration, a decreased cut-off in Q(50) of 36.5 units is suggested.

3.5 Fitting mixture distributions

Mixture distributions have been fitted to all 20 of the new samples (4 pure samples and 16 mixtures). Examples of the output are shown in Figures 7 to 10. The graphs are probability distributions for the five parameters of interest, ie the proportion of adulteration and the parameters (mean and sd) of each of the two components in the mixture. The figures show the estimated probability distributions of the proportion of admixture. Each figure shows a

Figure 7

Probability distribution of measured non-Basmati rice content in known mixtures.

Basmati 1 + non-Basmati 1

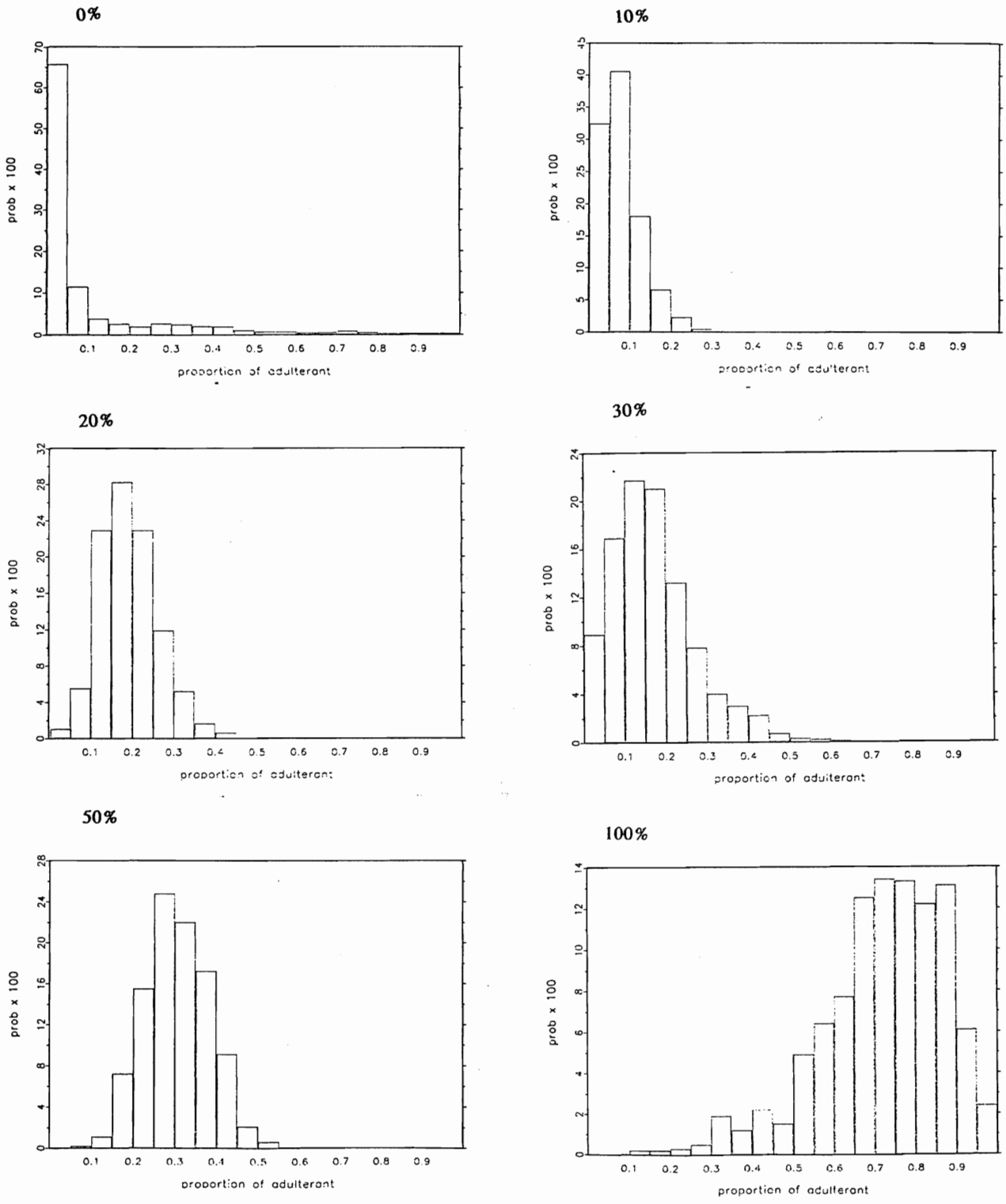


Figure 8

Probability distribution of measured non-Basmati rice content in known mixtures.

Basmati 1 + non-Basmati 2

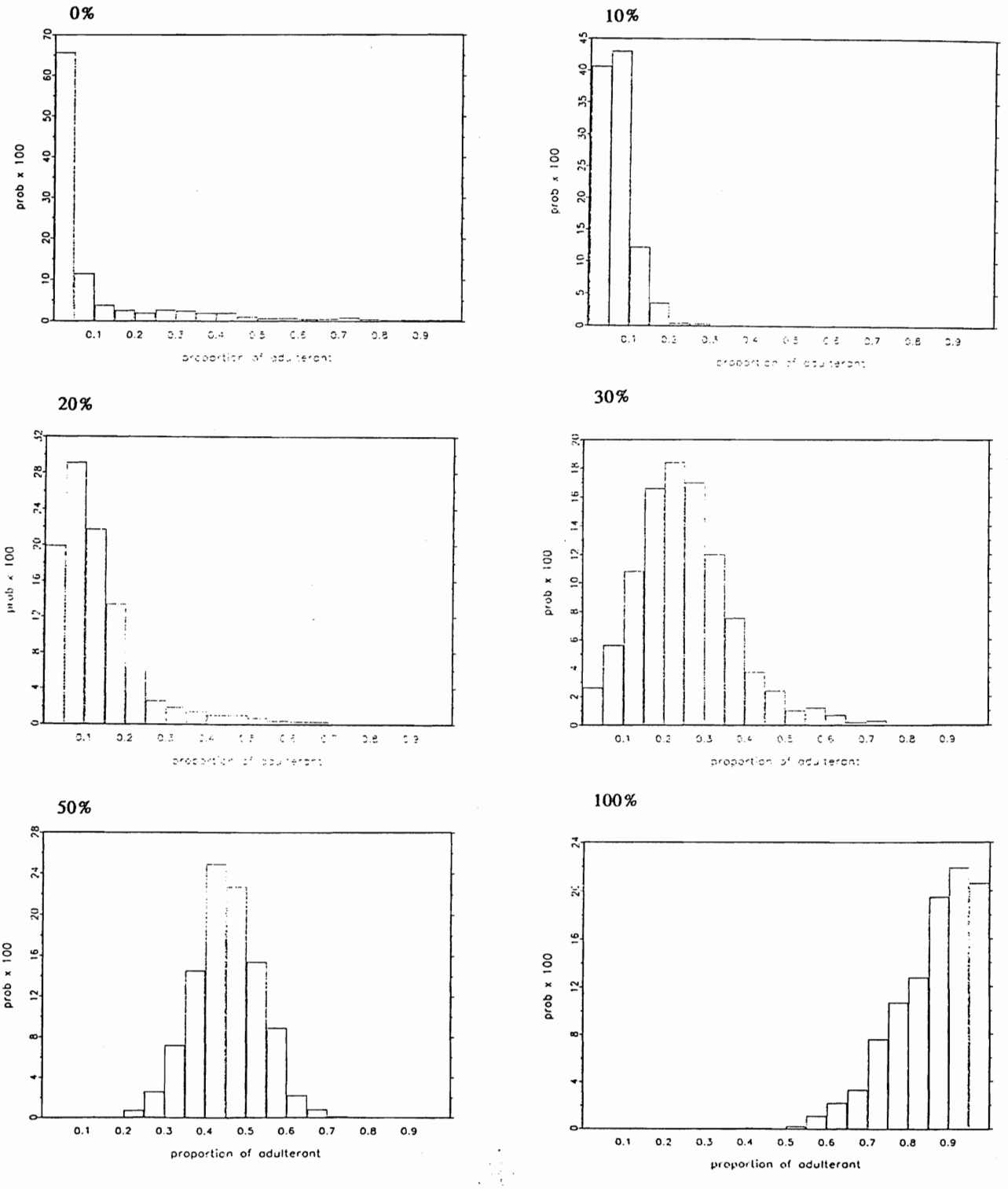


Figure 9

Probability distribution of measured non-Basmati rice content in known mixtures.

Basmati 2 + non-Basmati 1

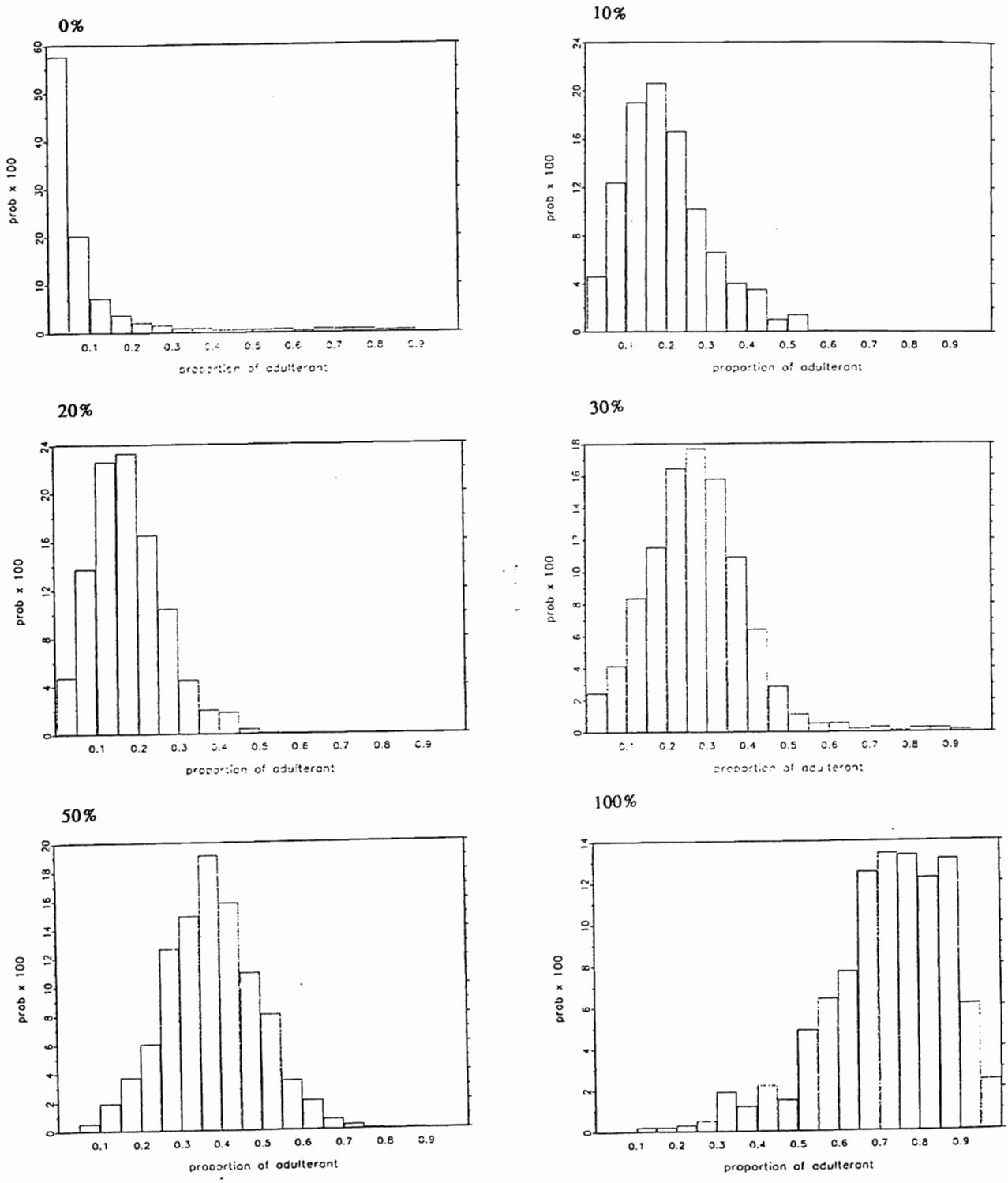


Figure 10

Probability distribution of measured non-Basmati rice content in known mixtures.

Basmati 2 + non-Basmati 2

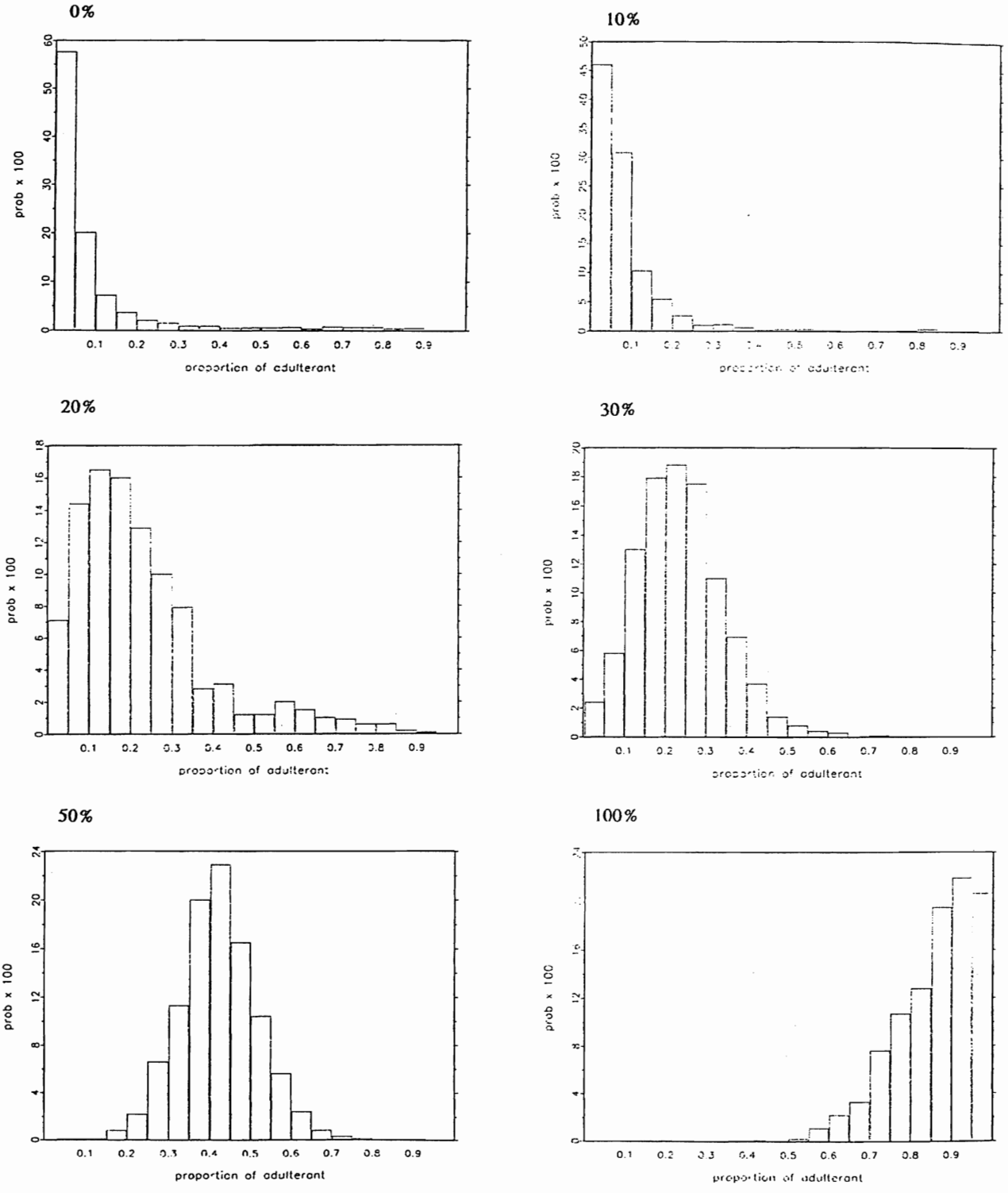
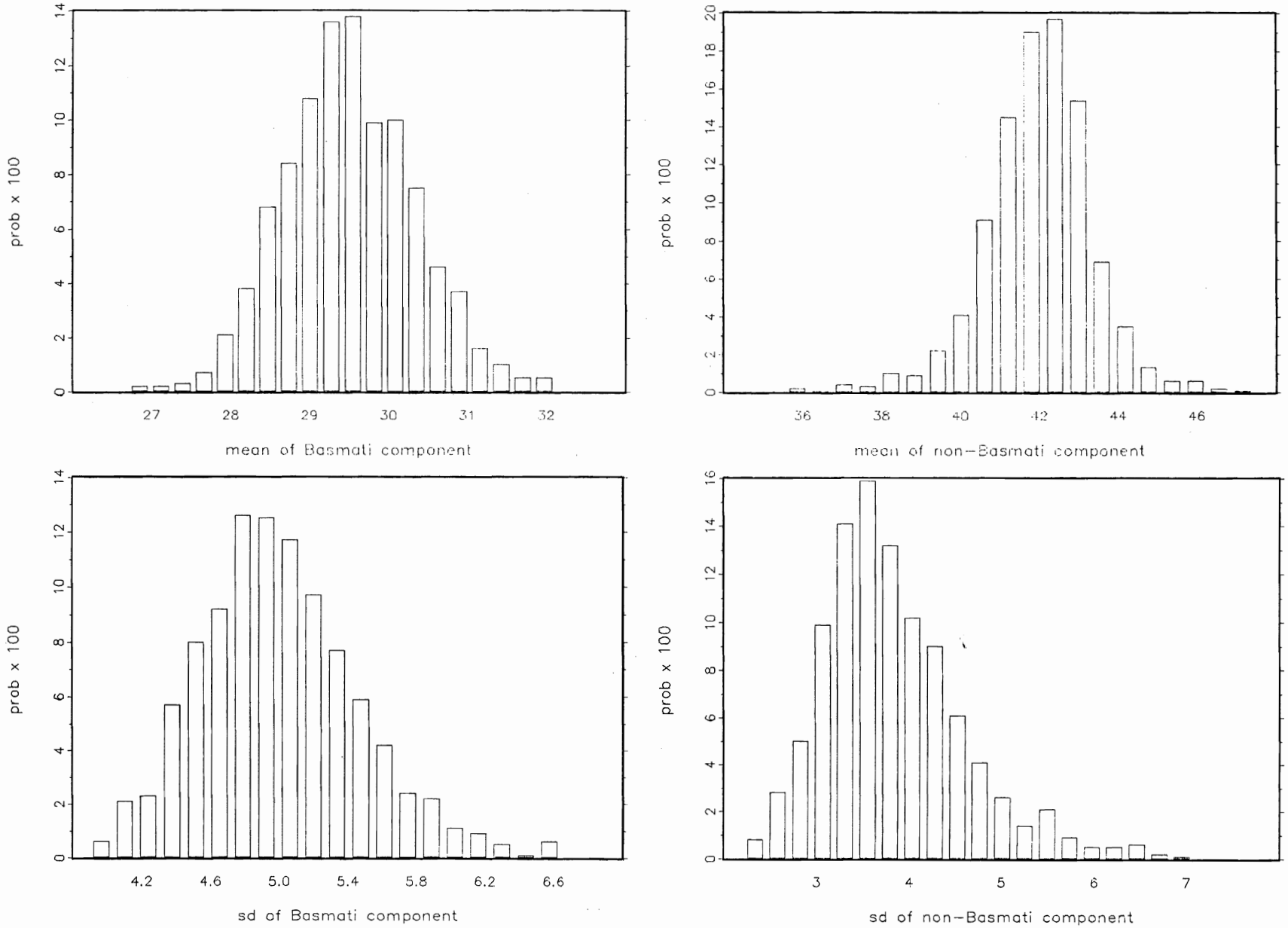


Figure 11

Probability distributions of the estimated mean and standard deviation RP3 values for each constituent of a real mixture (that of Figure 7; 20% admixture)



different pairing of a Basmati (#1 or #2 - see section 2.6) and non-Basmati sample (#1 or #2). The 6 plots within each figure correspond to the different actual levels of admixture used (0% - i.e. pure Basmati, 10%, 20%, 30%, 50%, 100% - i.e. pure non-Basmati). The samples used in the mixtures are indicated on the figures.

Although the relative sample proportion, as shown in Figures 7 to 10, is the parameter of greatest direct relevance to the quantification of admixture, the fitting procedure also generates probability distributions describing the likely nature (i.e. mean and s.d. of RP3) of each of the 2 components of the mixture. An example of these for the case of 80% Basmati 1 + 20% non-Basmati 1 is shown in Figure 11.

Although 500 grains have been measured, the results presented in Figures 7 to 11 are for random 50-grain subsamples drawn from these datasets, to provide consistency with the size of the training samples (46 grains or fewer) upon which the distributional assumptions are based. Caution should be exercised in extrapolating the training samples to 500 grains. Also, for 500 grain test samples, the sampling errors are sufficiently reduced for deviations of the pure sample RP3 distributions from normality to be detectable. Since the fitted distributions are assumed to be normal, such deviations may be wrongly attributed to the presence of mixtures and give inaccurate estimates of the true level of admixture.

In all cases, the results for 50-grain subsamples are consistent with the known composition of the sample, but there is considerable uncertainty. Some of this uncertainty is inherent in the problem; some is due to the small sample sizes, and may be reduced by further development of the method to work on larger samples.

4. DISCUSSION AND CONCLUSIONS

Previous studies (Osborne *et al.*, 1993b, Whitworth *et al.*, 1995) have shown image analysis to be capable of detecting substitution of Basmati rice with most other types of rice, on the basis of measurements made of as few as 50 cooked grains. However, the detection of admixture to Basmati rice is also an important aspect of quality control, and of guaranteeing authenticity. This more complex problem presents additional difficulties, because of the overlap of the within-sample distributions of measurements, necessitating larger sample sizes and more sophisticated statistical analyses.

Although generally successful, drifts had been observed in the measurements previously generated. New software and hardware have been adopted in the hope of reducing their likelihood. Further, new calibration procedures have been introduced to detect such drifts should they occur, and to facilitate transferability of the measurement to other comparable instruments. The optimal cooking time to achieve good discrimination of rice types has been reviewed, and previous practice found to be acceptable. Such studies have also allowed the sensitivity of results to the cooking time to be quantified, identifying the need to control this to within 30 s.

Two approaches to the statistical analysis of mixtures have been used in this project. The simpler one gives an indication of when Basmati samples contain admixture of other types. It provides a pass/fail decision based on a single discriminant function, and seems capable of indicating admixture at or above a level of about 20%, when a typical cut-off value of this function is applied. The performance of this test has been assessed on a limited number of real mixtures created from within the database. When applied to other samples, it is conceivable that false positives or false negatives may occasionally arise, particularly when the required level of detection is small. By adjustment of the cut-off, and depending on the required level of detection, it is possible to minimise the occurrence of the more undesirable class of false result, at the expense of the other. For example, for enforcement purposes, the failure to identify some cases of adulteration might be considered more acceptable than the accusation of innocent parties. Conversely, the interest of traders is the guaranteed authenticity of the products they offer for sale; such traders will not wish to purchase any

material which appears suspicious, even though this may involve the rejection of some genuine product. The reliable choice of cut-offs to meet these various criteria will depend on extension of the database to include further known reference materials. The advantage of the above statistical approach is its simplicity to calculate and use. However, being a simple pass/fail test, it does not quantify the uncertainty involved, or give a direct measure of the level of admixture.

The more sophisticated approach involves the statistical fitting of a weighted sum of two normal distributions to the measured data, and does quantify the uncertainty, giving an estimate of the most probable level of addition. It also characterises the individual constituents of a mixture of two rice types. The drawback with this approach is that it involves several assumptions (some of which may not be sufficiently justifiable for legal purposes) and depends on a substantial computer program, thus obscuring to the non-specialised user the basis on which the test operates. Due to the limited number of measurements made of the database reference samples, the method has thus far only been trained on the basis of 46 grains per sample, and it is uncertain whether extrapolation to 500 grains, on the assumption that their RP3 values are normally distributed, is valid. The test performs satisfactorily on test samples of 50 grains (although sampling errors at this level yield a high measure of uncertainty). Further statistical work is being undertaken (Fearn, 1996) to assess the validity of extrapolation, but ideally, larger reference sample datasets would be desirable for accurate training.

In summary, a method has been developed and documented which is capable of detecting likely addition of non-Basmati to Basmati rice at a level of typically 20%, depending on the acceptable level of false positives or negatives. The method involves measurement of 500-grain samples, which are statistically compared with a database of 46-grain measurements of 95 reference samples. The most sophisticated analysis procedure developed also provides a quantitative estimate of the level of addition, and of the uncertainty in this value. At present, the latter method has only been validated for 50 grain test samples. Further statistical analysis or measurement of reference samples would be necessary to allow it to be confidently applied to 500-grain samples which, if done, would reduce the uncertainty.

Future work

It is considered that the test described here is not yet suitable as the basis for prosecutions, due to the assumptions made in the data analysis, and to the limited scope and in many cases unproven purity of the reference database samples. However, it is considered that the test would provide useful assistance in purchasing decisions. It would also be suitable for survey work, assuming that the detection level is considered useful to the aims of such work. If the maximum benefit is to be obtained from any such surveillance activity, it is recommended that larger (i.e. 500-grain) quantities of reference samples should first be measured. This would allow the cut-off value of the simple statistic to be chosen and interpreted with greater confidence. Furthermore, it would enable the more complex, quantitative method to be validated for application to 500-grain test samples, for which a reduced uncertainty in the estimated proportion of admixture would be achievable. If such extension of the database were contemplated, it would be an appropriate moment to source additional samples of guaranteed provenance and purity.

REFERENCES

American Association of Cereal Chemists (1995) Determination of the pasting properties of rice with the Rapid Visco-Analyser. AACC Method 61-02.

Fearn, T. (1996) Detecting adulteration of Basmati rice by the analysis of mixtures. Research Report 169, Department of Statistical Science, University College London. (In press)

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995) Bayesian Data Analysis, Chapman & Hall: London.

Osborne, B.G., Mertens, B., Thompson, M. and Fearn, T. (1993a) The authentication of Basmati rice using near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 1, 77-83.

Osborne, B.G., Whitworth, M.B., Greenwell, P., Mertens, B., Thompson, M. and Fearn, T. (1993b) The authentication of Basmati rice using near infrared transmittance spectroscopy. Final Report on a Project for the Ministry of Agriculture, Fisheries and Food for the year ending 31 March 1993.

Scanlon, M.G., Pritchard, P.E. and Osborne, B.G. (1991) Determination of the authenticity of rice varieties. Report on Project N1662 for the Ministry of Agriculture, Fisheries and Food.

Whitworth, M.B., Greenwell, P., Osborne, B.G. and Fearn, T. (1995) Physical Techniques for Establishing the Authenticity of Rice. Final report on Project ANO 634 for the Ministry of Agriculture, Fisheries and Food.

ACKNOWLEDGEMENTS

The authors would like to thank MAFF for funding this work, and the Natural Resources Institute, USA Rice Council, CIRAD, Ente Nazionale Risi, Indian Directorate of Rice Research, Stevens and Brotherton Ltd., Whitworths Ltd., and GAFTA for assistance in collecting samples and for advice on their relevance.

APPENDIX

Procedure for Analysis of Basmati Rice

1. Title Determination of Basmati Rice Authenticity by Image Analysis

Note: This procedure describes the specific implementation of the method as used by the authors for the purposes of this project. Variations in detail may be permissible, but care should be exercised to ensure that results are unaffected (see section 2.7 of main report).

2. Scope

This method is applicable to samples of non-parboiled, white, milled rice. It should not be regarded as definitive, since it is insensitive to some types of non-Basmati rice.

3. Principle

When cooked, grains of Basmati rice typically have a more elongated shape than grains of other rice types. Grain images viewed in silhouette against a light box are digitised for computer analysis. They are then measured and each grain is characterised by a variable representing its degree of conformity to the size and shape typical of Basmati. The degree of conformity of a population of such values to that typical of a pure Basmati sample is an indication of the authenticity of the sample.

4. Apparatus

4.1 PC computer with installed Sprynt image processing board, and Semper software (Synoptics Ltd, Cambridge)

4.2 Boiling water bath (1kW power, 210mm diameter x 125mm height)

4.3 250 ml Pyrex beaker

4.4 Timer

4.5 Thermometer capable of measuring up to 100°C to an accuracy of at least 1°C.

4.6 Bung for beaker with hole for thermometer. If the bung is tight fitting, a second hole also should be provided for relief of steam pressure.

4.7 Retort stand and clamp suitable for holding beaker by its rim.

4.8 JVC colour video camera model TK870E with 35 mm lens

4.9 Light box (Model QUP/A45C DVV Viewboxes, 2 x 8W lamps)

4.10 Acetate film (as used for overhead projector transparencies).

4.11 Kaiser RS3 copy stand

4.12 Cardboard or other black cover for copy stand and camera.

4.13 Paper towels ('blue roll')

4.14 Blunt ended tweezers

4.15 Plastic kettle (to be reserved for use with distilled water only)

4.16 Gloves or towel suitable for safe handling of beaker of boiling water

4.17 Clear plastic ruler

5. Reagents

5.1 Distilled water

6. Procedure

The Synoptics Semper software runs in windows on a PC computer. For this procedure, library programs have been developed within Semper, appropriate for use with a Sprynt card.

Initially, an image of a light box is stored. Subsequently, cooked grains are placed on the lightbox, their images are grabbed, and are normalised by the original image. Regions darker than a fixed threshold value are identified as grains and measured. The measurements are converted from pixel units to true dimensions according to a scaling factor established by measurement of an image of a ruler. The dimensions of grains in a test sample are compared with distributions previously established for known samples to assess their degree of conformity with the Basmati type, and to determine the authenticity of the sample.

The cooking procedure (6.1) includes a delay of 10 minutes (6.1.6 - 6.1.7) while the rice cooks. During this time, it should be possible to perform the calibration procedure of 6.2, unless the system is being set up for the first time, in which case it may be necessary to carry out 6.2 before 6.1. By the time the rice has cooked, calibration must have been completed, so that the rice can be analysed without delay. During analysis, if any delay occurs of more than one or two minutes, the remaining grains should be rejected (since they dry out and shrink with time), and a fresh set of grains cooked and analysed if insufficient results have been accumulated.

In some instances, it may be necessary to stop the program, which can be achieved by pressing the *Escape* key. If it is required to re-run the program, you should first type the commands:

Deassign device 5 [*Ctrl* + *Enter*]
Delete 1,99 [*Ctrl* + *Enter*]

6.1 Cooking of Rice

The apparatus used to boil rice is shown in figure 1.

6.1.1 Fill the water bath with tap water and bring it to the boil.

6.1.2 Boil some distilled water in the kettle.

6.1.3 Ensure that the beaker has been washed. Mount it from the retort stand, gripped by a clamp at its rim. Suspend it through a circular hole in the lid of the water bath, such that the external water is level with at least the 100 ml graduation. Fill the beaker with about 200 ml of boiling distilled water from the kettle. Place the bung in the mouth of the beaker and suspend the thermometer through it such that the bulb is entirely immersed within the water.

6.1.4 Wait until the temperature of the water reaches at least 96°C, as measured by the thermometer. It is not possible for it to reach 100°C because heat is transferred from the water bath by conduction through the beaker. This requires a temperature gradient, and so the interior of the beaker will always be at a lower temperature than the exterior, and the water inside will not boil.

6.1.5 Ensure that the rice sample is thoroughly mixed. Tip at least 500 grains onto a sheet of paper. Reject any grains which are obviously broken, to give a total of at least 500 intact grains. Representatively divide these into subsamples of no more than 250 grains each. A maximum of 250 grains should be cooked at a time, to enable them to be analysed sufficiently rapidly to prevent them drying out after cooking. Further subsamples should be cooked by subsequent repetition of 6.1.

6.1.6 Temporarily remove the bung from the beaker and tip the intact grains into the water. Take care to avoid scalding by steam. Replace the bung and start the timer.

6.1.7 After 10 minutes have elapsed on the timer, remove the beaker from the water bath using gloves or a cloth for protection against the heat. Remove the bung, and swill the water round in the beaker to dislodge grains which have become stuck to the bottom. Decant off some of the water if necessary. Dislodge any remaining stuck grains gently with tweezers. Decant off the water. Pour on further boiling distilled water and swill it around in the beaker. Decant this off. Add some cold distilled water (the beaker may now be safely handled, but the rim may remain hot). This will stop the cooking process, but will prevent the rice from drying out. It should take about 1 minute between removing the beaker from the water bath, and applying the cold water. It is important that this time, and the cooking time of 10 minutes should be controlled to within at least 30 s, since the effects of cooking time on the results are known to be significant for variations of greater than this.

6.1.8 Fold a paper towel in four and pour the rice onto it, ensuring that it becomes thoroughly soaked with the water. Excess water may be decanted off. The purpose of the wet paper towel is to prevent the rice grains drying out during the time required to analyse them, thus minimising shrinkage. It is

important to proceed promptly with the analysis, and to avoid interruptions. The towel should be folded over to cover grains awaiting analysis, thus further preventing drying. Grains which have become broken during cooking should be rejected, and not analysed.

6.2 Calibration and alignment of imaging system

6.2.1 Mount the camera above the lightbox on the copy stand, facing vertically down. Cover the surface of the lightbox with the acetate sheet. This has a rectangle drawn on it, which should face down (this prevents it being erased when the acetate is cleaned).

6.2.2 Run the application program *SEMPER* from *WINDOWS*, by double clicking on its icon with the mouse. You will be prompted for the video source in use. If the camera is connected to the computer via a Spectra decoder, type 3 to select the *Spectra-rgb* source. If it is connected directly to the computer, type 1 to select the *JVC-rgb* source.

6.2.3 Type *LIB RICE5* at the command prompt and press the *Ctrl* and *Enter* keys together to execute the command. You will be prompted for a file number. Enter a value, (*nnn*) in the range 1 to 999. After analysis, the data will be stored in a file named *RICE.nnn*. If this file already exists, you will be prompted for an alternative number.

6.2.4 A live colour image will be displayed on one of the monitors, with bands in false colour representing contours of constant image brightness. The computer will prompt for a background image to be presented. Adjust the height of the camera on the post of the copy stand, and focus it on the surface of the lightbox. A rice grain placed on the box may help in judging the focus.

6.2.5 Adjust the camera aperture until the image brightness is intermediate between the magenta and royal blue contours, and such that the magenta contour has almost vanished (typically a small region towards the centre of the image).

6.2.6 Position the acetate sheet and lightbox such that the measurement area marked on screen as a red rectangle lies just within the image of the rectangle on the sheet. Tape the sheet in position at its corners.

6.2.7 Clean the surface of the acetate sheet with a damp paper towel, such that no dark specks are present within the rectangular analysis region. Click the left button of the mouse to store the image as the background.

6.2.8 If no previous calibration for image magnification has been carried out, the computer will require this to be done now. It will present a live image, and prompt for a ruler to be placed in the field of view. Place the ruler on the lightbox such that the millimetre scale is visible across a diagonal of the screen. If the magnification is correct, about 45mm of the scale should be

visible. Click the left mouse button to store the image. Select two points separated by at least 30 mm on the ruler using the cursor, clicking the left mouse button at each point. The computer will display the scaling factor on screen, as the size of one pixel in millimetres. If this is outside the range 0.0533 to 0.0541, exit the program to the SEMPER prompt, as described before 6.1, and return to step 6.2.2, readjusting the camera position in 6.2.4 until the value lies in this range. This range is intended to provide a variation of no more than about ± 1mm in the height of the camera.

6.2.9 The computer will ask whether you want to measure any calibrants. These are cylinders of nylon, of similar size and opacity to rice, and are measured to ensure consistency of results measured on different occasions. If any adjustments have been made to the system since calibrants were last measured, they should be re-measured now. If no calibrants have been measured since SEMPER was started, the computer will require them to be measured.

6.2.10 The computer will ask how many calibrants are to be measured. Enter the required number, which should be at least 3.

6.2.11 A live image will be presented and you will be prompted to present the calibration cylinders. Lay as many as possible of the chosen number lengthways on the light box such that they do not touch one another, and such that they do not overlap the edges of the video frame. Press the left mouse button to analyse the image.

6.2.12 Repeat step 6.2.11 with further calibration cylinders until the selected number of these have been analysed.

6.2.13 The computer will display the mean and standard deviation of the following measurements for the calibration cylinders:

Area, length, breadth, perimeter, Rice Parameter 3 (RP3)

These, and the results for each cylinder will also be written to the results file. Plot the value of RP3 for each cylinder on a control chart. If the result deviates from the reference value for a given cylinder by more than a defined figure, this is indicative of a fault, and corrective action must be taken before reliable results can be obtained.

6.3 Analysis of rice

6.3.1 Following 6.2.13, the computer is now ready to analyse rice grains, and will prompt for the number of grains to be analysed. Enter the required number. This should be at least 500. If insufficient grains have been cooked, these should be analysed now, and further grains should be cooked according to 6.1.

6.3.2 A live image will be presented, and you will be prompted to present some grains for measurement. Lay the grains on the light box with their

symmetry plane horizontal, such that they do not touch one another, and such that they do not overlap the edges of the video frame. Remove any broken grains. Press the left mouse button to analyse the image.

6.3.3 After analysis has started, the grains may be removed, and further grains placed on the light box to save time.

6.3.4 Repeat steps 6.3.1 to 6.3.3 until the previously specified number of grains have been measured. Repetition of 6.2 may be necessary to generate sufficient cooked grains.

6.4 Output of Results

6.4.1 When analysis of the specified number of grains has been achieved, the computer will also measure any remaining grains in the final image, and will display the total number of grains measured on screen. It will also display the mean and standard deviation of each of the measured parameters:

Area, length, breadth, perimeter, RP3.

Record these values.

6.4.2 The measurements for each grain will be saved to disc in the file C:\ROL\RICE\RAW\RICE.nnn, where nnn is the file number specified at the start of analysis. A histogram of the values of RP3 for each grain will be displayed on screen, and saved as a Semper PIC image to disc with the filename C:\ROL\RICE\PIC\RHIS.nnn. Record the location of the results file (which can first be renamed or moved to a different directory or disc if desired).

6.5 Recall of previously stored histograms

Previously stored histograms can be recalled for examination on screen, using the SEMPER library program RICEHIS, assuming that the histogram still has a PATH and name of the form C:\ROL\RICE\PIC\RHIS.nnn, where nnn is a number of 1 to 3 digits.

6.5.1 Run SEMPER from WINDOWS by double clicking on its icon, and select a video source as in 6.2.2. For the purposes of 6.5, the actual choice is irrelevant.

6.5.2 Type LIB RICEHIS at the command prompt to run the library program.

6.5.3 You will be prompted for the number of a histogram (nnn, as used in its filename). Enter the number, and the histogram will be loaded and displayed on screen.

6.6 Statistical analysis to determine likely presence of adulterants

6.6.1 Sort the RP3 values in the saved datafile into increasing order.

6.6.2 Calculate the 50th percentile, Q(50) (the median) and the 90th percentile, Q(90). Calculate the discriminant parameter, T = Q(90)-Q(50).

6.6.3 If $T \geq 8$ or $Q(50) \geq 36.5$, the sample has an RP3 distribution inconsistent with the pure samples in the database, which may be indicative of the presence of an adulterant. A fail result should therefore be assigned. If $T < 8$ and $Q(50) < 36.5$, the sample is deemed to pass the test.

7. Expression of Results

7.1 Number of grains analysed

7.2 Difference, T between 90th and 50th percentiles of the distribution of RP3 values

7.3 The median (50th percentile) RP3 value.

7.4 Pass or Fail decision (see 6.6.3).

7.5 The measurements stored on disc are in ASCII format, delimited by commas. The file includes:

- Lines of text with headings at various points in the file.
- Results for each calibration object measured, and a summary of the means and standard deviations of these values.
- Results for each rice grain measured, and a summary of the means and standard deviations of these values.

The measurements stored for each object are:
Length / mm, Breadth / mm, Perimeter / mm, Area / mm^2, and Rice Parameter 3.
The number of lines of measurements for individual calibrants or rice grains may vary, but the format of the file is otherwise consistent.

8. Performance Indicators

8.1 Limit of detection

Estimated as approximately 20% by number of grains. A lower limit can be achieved at the expense of an increased number of false positive results.

8.2 Limit of determination

Estimated at 20-30%, but accurate knowledge of this value will only be obtained after further evaluation.

8.3 Relationship of analyte to adulterant

The method involves measurement of cooked grain dimensions, which are related to the swelling properties of the rice, Basmati generally having a greater elongation than other types.

8.4 Sensitivity

Estimated at about 20%, but implementation of the proposed quantitative statistical analysis is likely to improve this.

8.5 Specificity

In previous analysis of reference samples, 3 of 74 non-Basmati samples tested were indistinguishable from Basmati in pure form. Further, a few other samples are only likely to be detectable at higher limits than those specified above.

8.6 Accuracy - trueness and precision

The method described does not provide a quantitative estimate of the level of admixture, but provides a pass / fail decision at a chosen level of admixture. Further testing will be required to determine the levels of false positives, and of false negatives at any chosen detection level. The accuracy of the proposed quantitative method is not yet known.

The precision of measurement of RP3 by image analysis is high compared to the between grain variation within a sample. Furthermore, it is safeguarded against bias by a prescribed calibration procedure, based on measurement of reference nylon cylinders.

9. Useful Features of the Image Analysis Method for Authenticity Testing

9.1 Cost

Estimated at £50 per sample if analysed in eonomical batches.

9.2 Scope for extension to other products / matrices

Measurement of grain dimensions can be used as a partial determinant of the variety and species of several cereals. However, in most cases, electrophoresis methods are more definitive. It is not known whether the difference in swelling of rice types under cooking is also exhibited by other cereals, but pulses are known to exhibit varietally dependent swelling characteristics.

9.3 Duration of analysis

Approximately 1 hour per sample

9.4 Practicability

Could be performed by the average technician after short training.

9.5 Availablity of equipment

The equipment required is readily available in the UK. However, certain computer programs and statistical procedures are proprietary.

10. Quality Control

Pure samples of Basmati, and known mixtures should be tested periodically to ensure that Basmati passes the test, and samples with admixture above the required detection level, fail.

It has been observed with previous computer systems, that the reading of RP3 can drift, probably due to blurring of the image resulting from noise in the digitising electronics. To guard against such effects, reference objects are measured as part of the described procedure, (6.2.9 to 6.2.13) at least every time that SEMPER is restarted. These are nylon cylinders of similar size and opacity to rice. If their results deviate from reference values by more than specified limits, this indicates that rice measurements could be expected to be subject to similar errors, and that corrective action should be taken.

11. Safety

Care should be taken when handling containers of boiling water to avoid burns from the containers or scalding by the water, or by steam. A cloth or thick gloves should be used for handling of potentially hot items.