Comparison of two methods for detecting and correcting systematic error in high-throughput screening data

Andrei Gagarin[1], Dmytro Kevorkov[1], Vladimir Makarenkov[2], and Pablo Zentilli[2]

[1] Laboratoire LaCIM, Université du Québec à Montréal,
C.P. 8888, Succ. Centre-Ville, Montréal (Québec), Canada, H3C 3P8
[2] Département d'Informatique, Université du Québec à Montréal,
C.P. 8888, Succ. Centre-Ville, Montréal (Québec), Canada, H3C 3P8

Abstract. High-throughput screening (HTS) is an efficient technological tool for drug discovery in the modern pharmaceutical industry. It consists of testing thousands of chemical compounds per day to select active ones. This process has many drawbacks that may result in missing a potential drug candidate or in selecting inactive compounds. We describe and compare two statistical methods for correcting systematic errors that may occur during HTS experiments. Namely, the collected HTS measurements and the hit selection procedure are corrected.

1 Introduction

High-throughput screening (HTS) is an effective technology that allows for screening thousands of chemical compounds a day. HTS provides a huge amount of experimental data and requires effective automatic procedures to select active compounds. At this stage, active compounds are called hits; they are preliminary candidates for future drugs. Hits obtained during primary screening are initial elements for the determination of activity, specificity, physiological and toxicological properties (secondary screening), and for the verification of structure-activity hypotheses (tertiary screening) (Heyse (2002)).

However, the presence of random and systematic errors has been recognized as one of the major hurdles for successful implementing HTS technologies (Kaul (2005)). HTS needs reliable data classification and quality control procedures. Several methods for quality control and correction of HTS data have been recently proposed in the scientific literature. See for example the papers of Zhang et al. (1999), Heyse (2002), Heuer et al. (2003), and Brideau et al. (2003).

There are several well-known sources of systematic error (Heuer et al. (2003)). They include reagents evaporation or decay of cells which usually show up as smooth trends in the plate mean or median values. Another typical error can be caused by the liquid handling or malfunctioning of pipettes. Usually this generates a localized deviation of expected values. A variation in the incubation time, a time drift in measuring different wells or different plates, and reader effects may appear as smooth attenuations of measurements over an assay. This kind of effects may have a significant influence on the selection process of active compounds. They can result in an underestimation (false negative hits) or overestimation (false positive hits) of the number of potential drug targets.

We have developed two methods to minimize the impact of systematic errors when analyzing HTS data. A systematic error can be defined as a systematic variability of the measured values along all plates of an assay. It can be detected, and its effect can be removed from raw data, by analyzing the background pattern of plates of the same assay (Kevorkov and Makarenkov (2005)). On the other hand, one can adjust the data variation at each well along the whole HTS assay to correct the traditional hit selection procedure (Makarenkov et al. (2006)). Methods described in Sections 3 and 4 originate from the two above-mentioned articles.

2 HTS procedure and classical hit selection

An HTS procedure consists of running samples (i.e. chemical compounds) arranged in 2-dimensional plates through an automated screening system that makes experimental measurements. Samples are located in wells. The plates are operated in sequence. Screened samples can be divided into active (i.e. hits) and inactive ones. Most of the samples are inactive, and the measured values for the active samples are significantly different from the inactive ones. In general, samples are assumed to be located in a random order, but it is not always the case in practice.

The mean values and standard deviations are calculated separately for each plate. To select hits in a particular plate, one usually takes the plate mean value μ and its standard deviation σ to identify samples whose values differ from the mean μ by at least $c\sigma$, where c is a preliminary chosen constant. For example, in the case of an inhibition assay, by choosing $c = 3$, we would select samples with the values lower than $\mu - 3\sigma$. This is the simplest and most widely-known method of hit selection. This method is applied on a plate-by-plate basis.

3 Correction by removing the evaluated background

This correction method is a short overview of the corresponding procedure of Kevorkov and Makarenkov (2005). To use it properly, we have to assume that all samples are randomly distributed over the plates and systematic error causes a repeatable influence on the measurements in all plates. Also, we have to assume that the majority of samples are inactive and that their average values measured for a large number of plates are similar. Therefore, the average variability of inactive samples is caused mainly by systematic error, and we can use them to compute the assay background. In the ideal case, the measurements background surface is a plane, but systematic errors can introduce local fluctuations in it. The background surface and hit distribution surface of an assay represent a collection of scalar values which are defined per well and are plotted as a function of the well coordinates in a 3-dimentional diagram.

An appropriate statistical analysis of experimental HTS data requires a preprocessing. This will ensure the meaningfulness and correctness of the background evaluation and hit selection procedures. Therefore, we use normalization by plate and exclude outliers from the computations. Keeping in mind the assumptions and pre-procession requirements, the main steps of this method can be outlined as follows:

- Normalization of experimental HTS data by plate,
- Elimination of outliers from the computation (optional),
- Topological analysis of the evaluated background,
- Elimination of systematic errors by subtracting the evaluated background surface from normalized raw data,
- Selection of hits in the corrected data.

3.1 Normalization

Plate mean values and standard deviations may vary from plate to plate. To compare and analyze the experimental data from different plates, we need first to normalize all measurements within each plate.

To do this, we use classical *mean centering and unit variance standardization* of the data. Specifically, to normalize the input measurements, we apply the following formula:

$$x'_i = \frac{x_i - \mu}{\sigma}, \tag{1}$$

where x_i, $i = 1, 2, \ldots, n$, is the input element value, x'_i, $i = 1, 2, \ldots, n$, is the normalized output element value, μ is the plate mean value, σ is the plate standard deviation, and n is the total number of elements (i.e. number of wells) in each plate. The output data will have the plate mean value $\mu' = 0$ and the plate standard deviation $\sigma' = 1$.

Another possibility discussed by Kevorkov and Makarenkov (2005) is to normalize all the plate values to a given interval. This normalization generally produces results similar to the described one.

3.2 Evaluated background

Systematic error is assumed to appear as a mean fluctuation over all plates. Therefore, an assay background can be defined as the mean of normalized plate measurements, i.e.:

$$z_i = \frac{1}{N} \sum_{j=1}^{N} x'_{i,j}, \tag{2}$$

where $x'_{i,j}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, N$, is the normalized value at well i of plate j, z_i is the background value at well i, and N is the total number of plates in the assay.

Clearly, Formula 2 is more meaningful for a large number of plates: in this case the values of inactive samples will compensate the outstanding values of hits. To make Formula 2 useful and more accurate for an assay with a small number of plates, one can exclude hits and outliers from the computations. Thus, the evaluated background will not be influenced by the outstanding values and will better depict systematic errors.

3.3 Subtraction of evaluated background

Analyzing the distribution of selected hits, we can tell whether any systematic error is present or not in the assay: hits should be more or less evenly distributed over all wells. Otherwise, the hit amounts vary substantially from one well to another indicating the presence of systematic errors.

Deviations of the evaluated background surface from the zero plane indicate an influence of systematic errors on the measured values. Therefore, it is possible to correct raw HTS data by subtracting the evaluated background, defined by Formula 2, from the normalized values of each plate, given in Formula 1. After that, we can reassess the background surface and hit distribution again.

4 Well correction method

This section is a concise description of the well correction approach presented in detail in Makarenkov et al. (2006). We have to make the assumptions stated in the previous section about input HTS data and positions of samples in wells. The main steps of the well correction method are the following:

- Normalization of all sample values by plate,
- Analysis of hit distribution in the raw data,
- Hit and outlier elimination (optional),
- Correction and normalization of samples by well,
- Normalization of all samples by plate,
- Selection of hits in the corrected data.

Similarly to the evaluated background approach, the normalization of all samples by plate is done here using the mean centering and unit variance standardization procedure described above. The hit distribution surface can be computed as a sum of selected hits by well along the whole assay. If this surface is significantly different from a plane, it implies the presence of systematic errors in the assay measurements. Excluding hits and outliers from the computation, we obtain the non-biased estimates for the mean values and standard deviations of inactive samples in plates.

4.1 Well correction technique

Once the data are plate-normalized, we can analyze their values at each particular well along the entire assay. The distribution of inactive measurements (i.e. excluding hits and outliers) along wells should be zero-mean centered if systematic error is absent in the dataset.

However, a real distribution of values by well can be substantially different from the ideal one. Such an example is shown in the article by Makarenkov et al. (2006). A deviation of the well mean values from zero indicates the presence of systematic errors. Experimental values along each well can have ascending and descending trends (Makarenkov et al. (2006)). These trends can be discovered using the linear least-squares approximation (e.g. the trends can be approximated by a straight line).

In the case of approximation by a straight line ($y = ax + b$), the line-trend is subtracted from or added to the initial values bringing the well mean value to zero (x denotes the plate number, and y is the plate-normalized value of the corresponding sample). For the analysis of large industrial assays, one can also use some non-linear functions for the approximation. On the other hand, an assay can be divided into intervals and a particular trend function characterizing each interval can be determined via an approximation. After that, the well normalization using the mean centering and unit variance standardization procedure is carried out. Finally, we normalize the well-corrected measurements in plates and reexamine the hit distribution surface.

5 Results and Conclusion

To compare the performances of the two methods described above, we have chosen an experimental assay of the HTS laboratory of McMaster University (`http://hts.mcmaster.ca/Competition_1.html`). These data consist of a screen of compounds that inhibit the *Escherichia coli* dihydrofolate reductase. The assay comprises 1250 plates. Each plate contains measurements for 80 compounds arranged in 8 rows and 10 columns. A description of the hit follow-up procedure for this HTS assay can be found in Elowe et al. (2005).

Table 1 shows that the proposed correction methods have slightly increased the number of selected hits. However, the standard deviation of selected hits by well and the χ-square values (obtained using the χ-square contingency test with α-parameter equal to 0.01; the null hypothesis, H_0, here is that the hit distribution surface is a constant plane surface) become smaller after the application of the correction procedures. Moreover, the well correction method allowed the corresponding hit distribution surface to pass the χ-square contingency test in both cases (using 2.5σ and 3σ thresholds for hit selection). Figure 1 shows that the hit distribution surfaces have become closer to planes after the application of the correction methods.

To demonstrate the effectiveness of the proposed correction procedures, we have also conducted simulations with random data. Thus, we have con-

	Raw data	Rem. backgr.	Well correct.	Raw data	Rem. backgr.	Well correct.
Hit selection threshold	3σ	3σ	3σ	2.5σ	2.5σ	2.5σ
Mean value of hits per well	3.06	3.13	3.08	6.93	6.93	7.03
Standard deviation	2.17	2.16	2.06	3.93	3.55	2.61
Min number of hits per well	0	0	0	1	2	2
Max number of hits per well	10	10	10	19	22	15
χ-square value	121.7	118	109.1	175.8	143.8	76.6
χ-square critical value	111.14	111.14	111.14	111.14	111.14	111.14
χ-square contingency H_0	No	No	Yes	No	No	Yes

Table 1. Results and statistics of the hit selection carried out for the raw, background removed (Rem. backgr.) and well-corrected (Well correct.) McMaster data.

sidered random measurements generated according to the standard normal distribution. The randomly generated dataset also consisted of 1250 plates having wells arranged in 8 rows and 10 columns. The initial data did not contain any hit. However, the traditional hit selection procedure has found 119 false positive hits in the random raw data using the 3σ threshold. The correction methods detected 117 (removed background) and 104 (well correction) false positive hits.

Then, we have randomly added 1% of hits to the raw random data. The hit values were randomly chosen from the range $[\mu - 3.5\sigma; \mu - 4.5\sigma]$, where μ denotes the mean value and σ denotes the standard deviation of the observed plate. After that, the data with hits were modified by adding the values $4c, 3c, 2c, c, 0, 0, -c, -2c, -3c$, and $-4c$ to the 1st, 2nd, ..., and 10th columns, respectively, thus simulating a systematic error in the assay, where the variable c was consequently taking values $0, \sigma/10, 2\sigma/10, \ldots$, and $5\sigma/10$. The value $c = 0$ does not create any systematic error, but bigger values of c increase systematic error proportionally to the standard deviation σ.

For each value of the noise coefficient c, hits were selected in the raw, background removed and well-corrected datasets using the 3σ threshold. The hit detection rate as well as the false positive and false negative rates were assessed. The hit detection rate was generally higher for both corrected datasets. Figure 2(a) shows that the background and well correction procedures successfully eliminated systematic error from the random data. Both methods were robust and showed similar results in terms of the hit detection rate. However, the well correction method systematically outperformed the background method in terms of the false positive hit rate (see Figure 2(b)).

In conclusion, we developed two statistical methods that can be used to refine the analysis of experimental HTS data and correct the hit selection procedure. Both methods are designed to minimize the impact of systematic error in raw HTS data and have been successfully tested on real and artificial datasets. Both methods allow one to bring the hit distribution surface closer to a plane surface. When systematic error was not present in the data, both correcting strategies did not deteriorate the results shown by the tra-

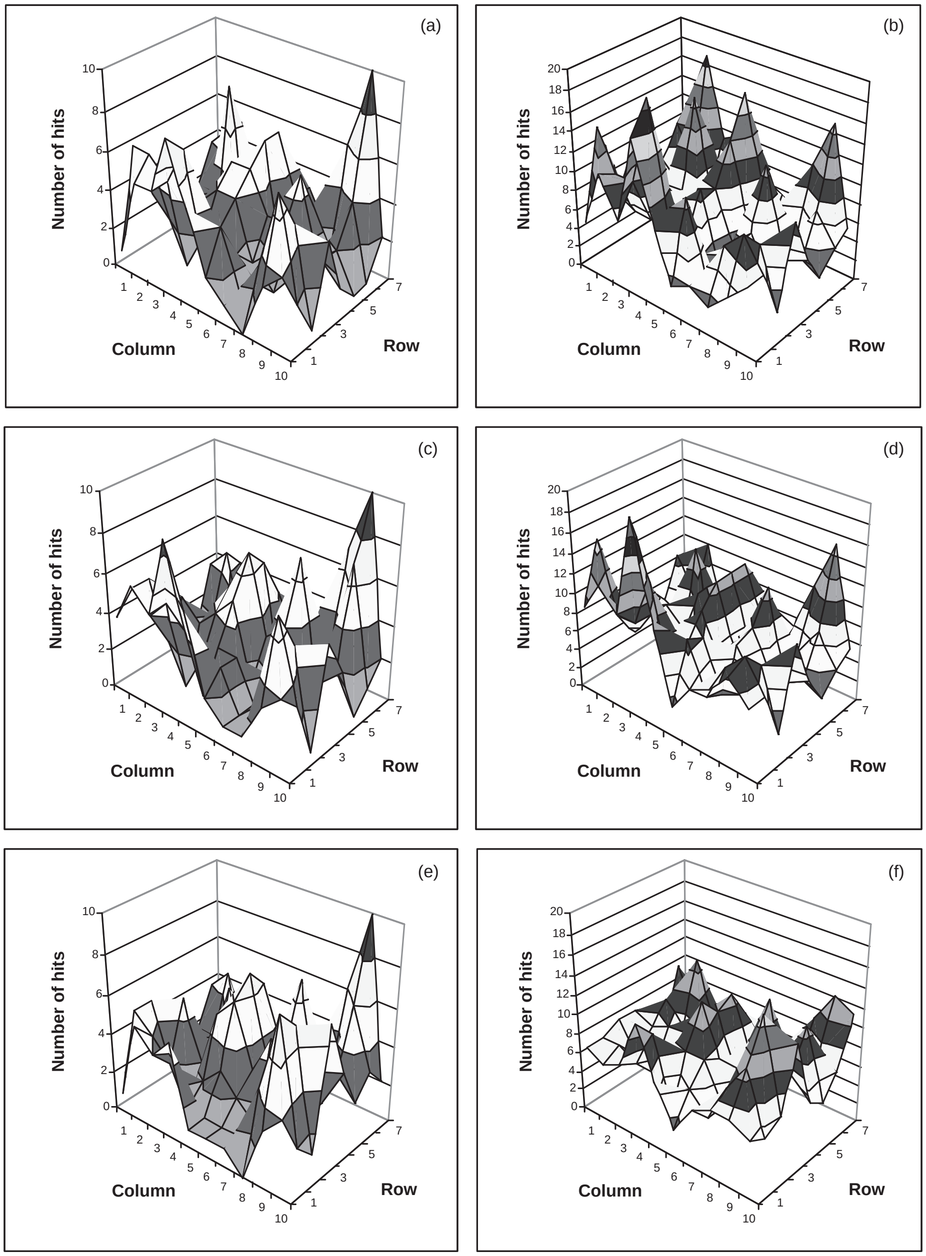

Fig. 1. Hit distribution surfaces computed for the 3σ and 2.5σ hit selection thresholds for the raw (a and b), background removed (c and d), and well-corrected (e and f) McMaster datasets.

ditional approach. Thus, their application does not introduce any bias into the observed data. During the simulations with random data, the well correction approach usually provided more accurate results than the algorithm proceeding by the removal of evaluated background.

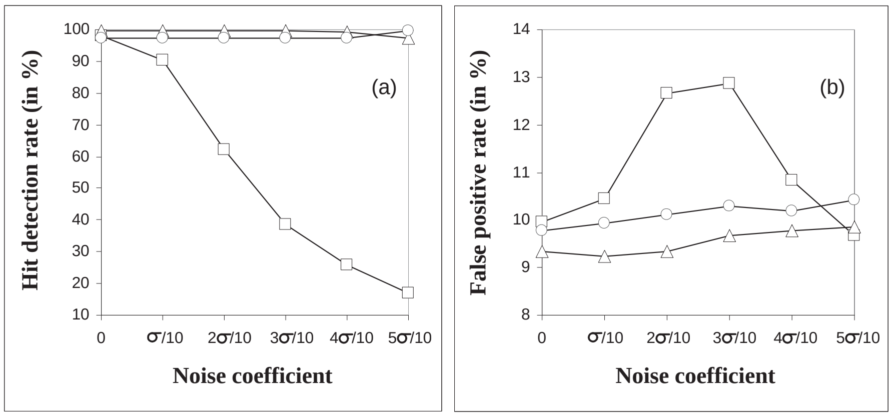

Fig. 2. Correct (a) and false positive (b) detection rates for the noisy random data obtained by the traditional hit selection procedure (denoted by □), the removed background (denoted by ○), and well correction (denoted by △) methods.

References

BRIDEAU, C., GUNTER, B., PIKOUNIS, W. and LIAW, A. (2003): Improved statistical methods for hit selection in high-throughput screening. *Journal of Biomolecular Screening, 8, 634-647.*

ELOWE, N.H., BLANCHARD, J.E., CECHETTO, J.D. and BROWN, E.D. (2005): Experimental screening of dihydrofolate reductase yields a "test set" of 50,000 small molecules for a computational data-mining and docking competition. *Journal of Biomolecular Screening, 10, 653-657.*

HEUER, C., HAENEL, T., PRAUSE, B. (2003): A novel approach for quality control and correction of HTS data based on artificial intelligence. *The Pharmaceutical Discovery & Development Report*. PharmaVentures Ltd. [Online].

HEYSE, S. (2002): Comprehensive analysis of high-throughput screening data. In: *Proceedings of SPIE 2002*, 4626, 535-547.

KAUL, A. (2005): The impact of sophisticated data analysis on the drug discovery process. *Business Briefing: Future Drug Discovery 2005*. [Online]

KEVORKOV, D. and MAKARENKOV, V. (2005): Statistical analysis of systematic errors in HTS. *Journal of Biomolecular Screening, 10, 557-567.*

MAKARENKOV, V., KEVORKOV, D., GAGARIN, A., ZENTILLI, P., MALO, N. and NADON, R. (2006): An efficient method for the detection and elimination of systematic error in high-throughput screening. Submitted.

ZHANG, J.H., CHUNG, T.D.Y. and OLDENBURG, K.R. (1999): A Simple Statistic Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of Biomolecular Screening, 4, 67-73.*