Data and text mining

Detecting and removing multiplicative spatial bias in high-throughput screening technologies

Iurie Caraus¹,², Bogdan Mazoure¹,², Robert Nadon²,³ and Vladimir Makarenkov¹,*

¹Département d’Informatique, Université du Québec à Montréal, Montréal, QC H3C-3P8, Canada, ²McGill University and Genome Quebec Innovation Centre, Montreal, QC H3A-0G1, Canada and ³Department of Human Genetics, McGill University, Montreal, QC H3A-1B1, Canada

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on January 25, 2017; revised on April 20, 2017; editorial decision on May 12, 2017; accepted on June 13, 2017

Abstract

Motivation: Considerable attention has been paid recently to improve data quality in high-throughput screening (HTS) and high-content screening (HCS) technologies widely used in drug development and chemical toxicity research. However, several environmentally- and procedurally-induced spatial biases in experimental HTS and HCS screens decrease measurement accuracy, leading to increased numbers of false positives and false negatives in hit selection. Although effective bias correction methods and software have been developed over the past decades, almost all of these tools have been designed to reduce the effect of additive bias only. Here, we address the case of multiplicative spatial bias.

Results: We introduce three new statistical methods meant to reduce multiplicative spatial bias in screening technologies. We assess the performance of the methods with synthetic and real data affected by multiplicative spatial bias, including comparisons with current bias correction methods. We also describe a wider data correction protocol that integrates methods for removing both assay and plate-specific spatial biases, which can be either additive or multiplicative.

Conclusions: The methods for removing multiplicative spatial bias and the data correction protocol are effective in detecting and cleaning experimental data generated by screening technologies. As our protocol is of a general nature, it can be used by researchers analyzing current or next-generation high-throughput screens.

Availability and implementation: The AssayCorrector program, implemented in R, is available on CRAN.

Contact: makarenkov.vladimir@uqam.ca

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Growing interest of the pharmaceutical industry has stimulated the development of several effective screening techniques such as high-throughput screening (HTS) and high-content screening (HCS). Modern HTS and HCS screening campaigns allow for examining hundreds of thousands of chemical compounds and generating gigabytes of experimental data (Lachmann et al., 2016). These data need to be extensively analyzed using appropriate statistical methods and protocols to detect potential drug candidates, called hits (Birmingham et al., 2009; Malo et al., 2006). A typical HTS assay is organized as a sequence of microtiter plates, featuring a grid of wells, which contain test samples. The most common plate formats consist of 96, 384, 1536 and 3456-well plates.
Spatial bias within plates (i.e. positional bias or systematic error) remains one of the major hurdles of experimental screening campaigns. It can be caused by a number of technical or environmental factors, including reader and pipette effects, liquid processing anomalies, unintended variations in compound concentration associated with agent evaporation, irregular changes in the incubation time, or temperature, lighting and air flow fluctuations (Heyse, 2002; Makarenkov et al., 2007). Spatial bias is evident as systematic under- or over-estimation of specific screen measurements (Kevorkov and Makarenkov, 2005). It is a constant source of false positives (i.e. inactive compounds incorrectly identified as hits) and false negatives (i.e. undetected active compounds) in screening technologies (Birmingham et al., 2009). Typically, spatial bias affects either compounds placed in the same well, row or column location over all plates of the assay (i.e. assay-specific error) or compounds from a specific row or column of a given plate (i.e. plate-specific error) (Dragiev et al., 2012). As we will show in this paper, spatial bias can also be of the additive or multiplicative nature.

Figure 1 illustrates an example of positional bias that affects the RNAi HIV inhibition assay screened at Pasteur Institute of Korea (Carralot et al., 2012). This cell-based assay uses HeLa P4 LTR-EGFP 2B4 cells, engineered to express the HIV cellular-entry receptors CD4 and CCR5. Both assay-specific (Fig. 1a) and plate-specific (Fig. 1b) spatial biases are present in this assay. Moreover, these biases exhibit opposite trends. On one hand, higher hit counts can be observed in the middle of the hit distribution surface (Fig. 1a; hit counts per well location are depicted), whereas lower hit counts can be observed at its edges (in rows A, O and P and in columns 23 and 24). On the other hand, the measurements in rows A, C and P and in columns 23 and 24 of Plate 7 (Fig. 1b) are systematically overestimated (hits correspond to the lowest measurements in inhibition assays).

Several data normalization and quality control techniques have been proposed to allow for an efficient evaluation and validation of HTS and HCS assays (Brideau et al., 2003; Carralot et al., 2012; Dragiev et al., 2011, 2012; Kevorkov and Makarenkov, 2005; Makarenkov et al., 2007; Malo et al., 2006; Murie et al., 2013; Shun et al., 2011; Zhang et al., 1999). The most popular data normalization methods used to compare experimental measurements across different plates of a given assay are Percent of Control (POC), Normalized Percent Inhibition (NPI), Z-score and robust Z-score (Birmingham et al., 2009; Malo et al., 2006). These methods, however, do not correct for spatial bias. A number of correction tools for the detection and removal of spatial bias from experimental screening data have been proposed (Caraus et al., 2015; Mpindi et al., 2015).

The popular B-score method (Brideau et al., 2003) relies on the median polish procedure (Tukey, 1977) in order to remove plate-specific spatial bias. Well Correction (Makarenkov et al., 2007) removes assay-specific bias using Z-score normalization and linear regression across well locations. R-score (Wu et al., 2008) is a plate-specific bias correction method which fits a robust linear model to experimental well measurements. Dragiev et al. (2012) presented the partial mean polish method intended for removing the additive plate-specific spatial bias from the data by correcting the biased measurements only (i.e. aPMP method). The SPAWN method (Murie et al., 2013) uses an iterative approach based on trimmed mean polishing to minimize row and column systematic effects within a given plate. Then, a well normalization step can be carried out to create a spatial bias template, which is used to correct assay-specific bias. However, almost all data correction methods, employed in screening technologies including B-score, R-score, aPMP and SPAWN, employed in screening technologies are designed to remove additive type of spatial bias. A notable exception is the diffusion model (Carralot et al., 2012), proposed to eliminate edge effects in RNAi HTS due to multiplicative bias.

The additive spatial bias model can be described by Equation (1):

\[
\bar{x}_{ip} = x_{ip} + R_{ip} + C_{ip} + e_{ip},
\]

whereas the multiplicative bias model by Equation (2):

\[
\bar{x}_{ip} = x_{ip} \times R_{ip} \times C_{ip} + e_{ip},
\]

where \( \bar{x}_{ip} \) is the resulting (biased) measurement value in well (\( i, p \)) of plate \( p \), \( x_{ip} \) is the original error-free measurement, \( R_{ip} \) is the bias...
affecting row \(i\) of plate \(p\), \(C_{ip}\) is the bias affecting column \(j\) of plate \(p\) and \(e_{ijp}\) is the random error in well \((i,j)\) of plate \(p\).

The most straightforward approach for removing multiplicative spatial bias in screening technologies consists of the use of a logarithmic transformation of raw measurements, followed by the application of one of the above-mentioned additive bias correction methods. However, a typical data preprocessing step in HTS consists of normalizing raw measurements (e.g. plate or well-wise using Z-score) prior to their correction and/or hit selection. Thus, the logarithmic transformation cannot be applied in many cases because the normalized measurements contain negative values. Furthermore, the multiplicative PMP method presented in Section 2.3 allows for minimizing multiplicative spatial bias by modifying only the biased measurements and keeping the corrected data on the same scale with the unbiased raw data. This property makes the data corrected by the proposed multiplicative PMP method directly comparable to those corrected by the additive PMP method (Dragiev et al., 2012).

This is further exploited in our general data processing algorithm presented in Section 2.5, where we show how to select the most appropriate bias correction model (additive or multiplicative) for the data at hand. Obviously, such a model selection algorithm could not be carried out if one of the two methods would work with the raw data, while the other used the log-transformed data.

Once spatial bias has been removed from the data, assay quality estimation and hit identification steps can be carried out (Birmingham et al., 2009). The most popular hit selection procedure identifies as hits the samples whose measurements are lower (for inhibition assays) or higher (for activation assays) than the selected threshold, where \(\mu\) is the mean, \(\sigma\) is the standard deviation of the target group of measurements, and \(c\) is a selected constant (often varying between 1 and 3).

In this article, we describe and compare three novel methods for removing the multiplicative spatial bias from experimental screening data. We then present a comprehensive bias correction protocol, which can be used to remove both additive and multiplicative spatial biases, as well as both assay and plate-specific biases. We apply the methods to real data to correct the data generated during a genome-wide siRNA screen aimed at studying HIV-host interactions (Carralot et al., 2012) and determine the dominant type of plate-specific spatial bias characterizing the four HTS screening categories (homogeneous, microorganism, cell-based and gene-expression) available in the ChemBank database (Seiler et al., 2008).

## 2 Materials and methods

### 2.1 Three new methods for correcting multiplicative bias

Here, we describe and compare three new methods for correcting multiplicative spatial bias in screening technologies. These methods are designed to remove plate-specific bias. Ideally, compound measurements and hit counts should be uniformly distributed over a given plate and the hit distribution surface of the assay. The first method, called NLMBE, solves a system of nonlinear algebraic equations in which the unknowns correspond to spatial biases, which affect specific rows and columns of a given plate. The second method, called multiplicative PMP (mPMP), is based on a multiplicative partial mean polynomial procedure in which the mean of each row and each column affected by spatial bias is adjusted iteratively with respect to the mean of the unbiased plate measurements. Rows and columns affected by spatial bias can be detected, for example, by the Mann–Whitney U test (Gibbons and Chakraborti, 2011; Wilcoxon, 1945). This information is required by both NLMBE and mPMP. The third method, called multiplicative B-score, is an adaptation of a 2-way median polish procedure (Tukey, 1977) to the case of multiplicative bias.

### 2.2 Non-linear multiplicative bias elimination (NLMBE)

The first method proceeds by solving a system of nonlinear equations in which the unknowns correspond to systematic errors affecting the measurements of a given plate. Let \(x_{ijp}\) be experimental measurements of plate \(p\), where \((i=1, \ldots, m)\) and \((j=1, \ldots, n)\), having \(m\) rows and \(n\) columns. Assume that some rows and/or columns of \(p\) contain spatial bias, as indicated for example by the Mann–Whitney U test. Let \(\mu\) be the mean (or median) of the measurements of plate \(p\) that do not contain spatial bias. The following system of nonlinear equations can be composed:

\[
R_p \times \left( \sum_{j=1}^{n} C_{jp} \times x_{ijp} / W_{ij} \right) = n \times \mu, \quad \text{for all } i = 1, \ldots, m, \quad (3)
\]

\[
C_{jp} \times \left( \sum_{i=1}^{m} R_p \times x_{ijp} / W_{ij} \right) = m \times \mu, \quad \text{for all } j = 1, \ldots, n, \quad (4)
\]

where \(R_p\) is the multiplicative bias affecting row \(i\) of plate \(p\), \(C_{jp}\) is the multiplicative bias affecting column \(j\) of \(p\) and \(W_{ij}\) is the systematic measurement offset in well \((i,j)\) across all plates of the assay. \(W_{ij}\) can be estimated directly from the assay background surface. This system of nonlinear equations is obtained by summing up all multiplicative bias terms corresponding to rows and columns of \(p\). It includes \(m + n\) equations and at most \(m + n\) unknowns (i.e. biases \(R_p\) and \(C_{jp}\) affecting the rows and columns of plate \(p\)). The values of \(R_p\) and \(C_{jp}\) that correspond to row \(i\) and column \(j\) not affected by bias are equal to 1. Since in practice only a few rows and columns per plate are affected by spatial bias, this system will have more equations than unknowns. Such systems can be solved, for example, by using the Levenberg–Marquardt (Moré, 1978) method based on nonlinear least-squares. Note that the Levenberg–Marquardt algorithm has a quadratic convergence rate. The corrected plate measurements \(\tilde{x}_{ijp}\) (Equation 2) can then be calculated taking into account the system’s solution, i.e. the obtained values of \(R_p\) \((i=1, \ldots, m)\) and \(C_{jp}\) \((j=1, \ldots, n)\).

### 2.3 Multiplicative PMP method (mPMP)

Our second method is based on an iterative procedure in which the mean of each row and each column affected by systematic error is gradually adjusted with respect to the mean of the plate measurements not affected by spatial bias. Assume that rows \(r_1, r_2, \ldots, r_k\) and columns \(c_1, c_2, \ldots, c_l\) of plate \(p\) contain multiplicative spatial bias. First, for any row \(i\) affected by spatial bias \((i=r_1, r_2, \ldots, r_k)\), we calculate:

\[
x_{ijp} = \mu_{i} \times x_{ijp} / (\mu_{i} \times W_{ij}), \quad \text{for all } j = 1, \ldots, n, \quad (5)
\]

and for any column \(j\) affected by systematic bias \((j=c_1, c_2, \ldots, c_l)\), we calculate:

\[
x_{ijp} = \mu_{j} \times x_{ijp} / (\mu_{j} \times W_{ij}), \quad \text{for all } i = 1, \ldots, m, \quad (6)
\]

where \(\mu_{i}\) is the mean of row \(i\), \(\mu_{j}\) is the mean of column \(j\), \(\mu\) is the mean of the plate’s measurements that are not affected by spatial bias and \(W_{ij}\) is the systematic measurement offset in well \((i,j)\) across all plates of the assay. When a large number of hits or outliers are expected, the means of the plate’s unbiased measurements, rows and
columns should be replaced by the corresponding medians in order to obtain more robust parameter estimates.

This iterative procedure should be repeated until a desired convergence threshold is reached. The time complexity of mPMP is $O(nml)$, where $n$ and $m$ are the plate dimensions and $l$ is the number of iterations required for convergence. In practice, this method converges after a few iterations. Importantly, mPMP is usually much faster than NLMBE.

The main advantages of the NLMBE and mPMP algorithms are that they modify only biased data and keep raw and corrected plate measurements on the same scale.

### 2.4 Multiplicative B-score method

We also present the multiplicative version of the well-known B-score algorithm (Brideau et al., 2003). The conventional (additive) B-score is a robust data correction procedure widely used in experimental screening technologies. Our multiplicative B-score transformation assumes the following bias model:

$$ \hat{x}_{ip} = \mu_p \times R_{ip} \times C_{ip} \times W_{ij}, \quad (7) $$

where $\hat{x}_{ip}$ is the estimated (biased) activity measurement in well $(ij)$ of plate $p$, $\mu_p$ is the average of plate $p$, $R_{ip}$ is the spatial bias affecting row $i$, $C_{ip}$ is the spatial bias affecting column $j$ and $W_{ij}$ is the systematic measurement offset in well $(ij)$. Our method is based on a 2-way median polish procedure (Tukey, 1977) in which subtractions are replaced by divisions in order to remove multiplicative spatial bias from all rows and all columns of $p$. The residual, $r_{ip}$, of the measurement in well $(ij)$ is defined as the difference between the raw measurement $x_{ip}$ and its fitted value $\hat{x}_{ip}$:

$$ r_{ip} = x_{ip} - \hat{x}_{ip}, \quad (8) $$

where $x_{ip}$ is the raw measurement in well $(ij)$ of plate $p$. Finally, the B-score is calculated as follows:

$$ B_{score} = \frac{\hat{r}_{ip}}{MAD_p}, \quad MAD_p = \text{median } |r_{ip} - \text{median } (r_{ip})|, \quad (9) $$

where $MAD_p$ is the adjusted median absolute deviation obtained from the residuals of plate $p$. The time complexity of the multiplicative B-score method is also $O(nml)$.

### 2.5 General data correction protocol

Here we present a complete bias correction protocol that can be used to remove both multiplicative and additive spatial biases, which can be assay or plate-specific.

First, *assay-specific bias* can be removed from a given assay by applying either the Well Correction (Makarenkov et al., 2007), based on Z-score normalization, or SPAWN (Murie et al., 2013), based on the robust Z-scores normalization, method. These methods normalize the measurements of specific well locations in which the presence of spatial bias has been detected (Dragiev et al., 2011). It is worth noting that the conventional Z-scores, when applied well-wise, allow for removing both additive and multiplicative spatial biases (Brideau et al., 2003). Following these normalizations, some data will become negative, thus making the use of the logarithmic transformation impossible.

Second, we propose the following algorithm to detect and remove *plate-specific spatial bias*. This algorithm should be applied in turn on all plates of a given assay. Carry out the Mann-Whitney $U$ test on each plate of the assay in order to detect biased rows and columns. This test will allow us to compare the sum of ranks of a given row or column to the sum of ranks of the rest of the plate’s measurements.

If (spatial bias is detected in some row(s) or column(s) of the plate), then:

1. Use the additive PMP method (Dragiev et al., 2012) to correct the plate’s measurements.
2. Use the multiplicative PMP method to correct the plate’s measurements.
3. Carry out the Kolmogorov-Smirnov two-sample test in order to compare first the distributions of unbiased measurements with those corrected by the additive PMP, and, second, the distributions of unbiased measurements with those corrected by the multiplicative PMP. Compute the $P$-values associated with these two corrections.
4. If (either the additive or multiplicative $P$-value from the previous step is larger than the selected significance level $\alpha$), then apply the correction algorithm that yields the highest $P$-value (i.e. additive or multiplicative PMP) to remove spatial bias from the plate’s data; otherwise, the bias model for this plate is undetermined.

Here, the Mann-Whitney $U$ test is applied on a plate-by-plate basis. The measurements of a considered row or column of a given plate compose the first vector used in the Mann-Whitney $U$ test and the rest of the plate’s data compose the second vector used in this test. If enough evidence for the presence of spatial bias, expressed through the test’s $P$-value, is obtained, the corresponding row or column is flagged as biased and removed from the computation. If no biased rows or columns have been found at the current iteration, the procedure is stopped. In our study, the maximum number of iterations allowed by the algorithm was limited to 50% of the total number of rows (when the presence of bias in rows was examined) and 50% of the total number of columns (when the presence of bias in columns was examined) of a given plate.

It is worth noting that when the plate’s background estimation is close to zero, the Mann-Whitney $U$ test should not detect any biased row or column within a given plate. We compared the results of the Mann-Whitney $U$ test to those of the $t$-test used by Dragiev et al. (2012) in terms of spatial bias detection, and found that the Mann-Whitney $U$ test is more robust in this context and thus better suited for spatial bias detection purposes in screening technologies. The main advantages of the Mann-Whitney $U$ test compared to the $t$-test are that it does not make any distributional assumptions and is more robust to outliers.

### 3 Results

To evaluate the performance of the three spatial bias correction methods, we carried out simulations with artificially generated screening data. Afterwards, we compared the most successful of them, multiplicative PMP, to the existing data correction techniques such as B-score (Brideau et al., 2003) and diffusion model (Carralot et al., 2012; Ogier and Dorval, 2012) considering RNAi HIV inhibition assay, screened at Pasteur Institute of Korea (Carralot et al., 2012). We also used our general data correction protocol to assess the extent of plate-specific bias across the data corresponding to different HTS categories available in ChemBank.

#### 3.1 Simulation study

Our simulation study was conducted using randomly generated 1000-plate assays. The three considered plate sizes were as follows:
96-well plates (8 rows x 12 columns), 384-well plates (16 rows x 24 columns) and 1536-well plates (32 rows x 48 columns). The values of inactive activity measurements followed a normal distribution with parameters ($\mu = 7.344$ and $SD = 1$), where $\mu$ and $SD$ were the mean and the standard deviation of the plate's measurements. Active compounds (hits) were randomly added to the plates to obtain assays with the following hit percentages: 0%, 0.5%, 1%, 2%, 3%, 4% and 5%. Hit locations on each plate were randomly chosen following a uniform distribution. Hit measurements were generated via sampling from a normal distribution with parameters $\sim N(\mu, 1.67SD, SD)$. A multiplicative spatial bias was randomly assigned to some rows and columns of all plates of the assay. The bias value was selected following a normal distribution with parameters $\sim N(1, C)$, with $C$ equal to 0, 0.1SD, 0.2SD, 0.3SD, 0.4SD and 0.5SD.

In our simulations, the number of rows and columns of a given plate which could be affected by systematic bias was limited to a maximum of 4 rows and 4 columns for 96 and 384-well plates, and a maximum of 5 row and 5 columns for 1536-well plates. A small random noise was also added to both hit and non-hit measurements on all plates. The noise values followed a normal distribution with parameters $\sim N(0, 0.5SD)$. The biased measurements of a given plate were generated using Equation (2). The four following data correction methods were tested in our simulations: No Correction, NLMBE, mPMP and multiplicative B-score. The Mann-Whitney $U$ test was used in the NLMBE and mPMP methods to identify the rows and columns affected by spatial bias. The hits were chosen globally across all assays by using the hit selection threshold of $\mu_{hit} - 1.67SD_{hit}$ (i.e. all measurements lower than $\mu_{hit} - 1.67SD_{hit}$ were chosen as hits, where $\mu_{hit}$ and $SD_{hit}$ were the mean and the standard deviation of a given assay after the addition of both hits and spatial bias). The performance of our data preprocessing techniques was assessed by measuring the total number of false positives (FP) and false negatives (FN) as well as by computing the hit detection rate (i.e. the true positive rate) for all methods.

Two series of experiments were conducted, by varying either the hit percentage or the bias level. In the first set of experiments, 1000 different assays with a fixed standard deviation of bias, equal to 0.3SD, and the hit percentage rate varying from 0% to 5% were generated for each plate size (there are no true positives for the case of 0% of hits; Figs 2–4a and b). In the second series of experiments, 1000 different assays with the fixed hit percentage of 1% and the standard deviation of bias varying from 0 to 0.5SD were generated for each plate (Figs 2–4c and d). Figures 2, 3 and 4 present the average results generated by the four compared methods for the 96-well, 384-well and 1536-well plates, respectively. The results of our simulations suggest that the NLMBE and mPMP methods clearly outperformed the No Correction procedure regardless of plate size, hit percentage and spatial bias variance (see Figs 2–4). Moreover, NLMBE and mPMP usually outperformed the multiplicative B-score method. This trend is most noticeable with 96-well plates. The results of multiplicative B-score improved with the increase in plate size. It is worth noting that the multiplicative B-score method was very prone to generating false positives, especially for 96 and 384-well plates (see also Mpindi et al. (2015) for a discussion on drawbacks of this method). The NLMBE method generally achieved slightly better performances than mPMP in terms of detection rate (i.e. true positive rate). However, mPMP was slightly better than NLMBE in terms of the combined false positive and false negative rate. Considering that mPMP converges much faster than NLMBE, it slightly better performances than mPMP in terms of detection rate (i.e. true positive rate for inhibition assays). These changes allowed our new methods, mainly mPMP and NLMBE, to achieve good correction and hit selection results, especially for large plates (Fig. 5e and f).

Moreover, we also conducted simulations with higher hit rates, i.e. up to 20% (see Fig. 5), which may occur in secondary screening. As in this case the plate’s (the row’s or the column’s) mean can be heavily affected by outliers, and because the values of hits can be viewed as outliers, we used the median instead of the mean in all our calculations within the mPMP, NLMBE and multiplicative B-score methods when working with secondary screening data. We also used the median instead of the mean in the formula defining the hit selection threshold (i.e. $Hits \text{ values} \leq \text{Median} - C \times SD$ for inhibition assays). These changes allowed our new methods, mainly mPMP and NLMBE, to achieve good correction and hit selection results, especially for large plates (Fig. 5e and f).
We also carried out a simulation involving different plate layouts and the proposed mPMP bias correction method. Precisely, we compared hit detection results obtained for 384-well plates (a) without controls (i.e. all of the plate’s wells comprised regular screening samples), (b) with the control layout corresponding to Figure 1a in Mpindi et al. (2015) (i.e. layout based on placing controls in column 1 and 24) and (c) with the control layout corresponding to Figure 1b in Mpindi et al. (2015) (i.e. layout based on randomly scattering controls across the entire plate). The hit rate in this simulation varied from 1 to 20%. The positions of controls (but not their values) were taken into account in this simulation during the computation of the method’s parameters. Even though superior results were obtained for plates with no controls, followed by plates with the scattered control layout, and, finally, by plates with controls located in the first and last columns, the results of this simulation, presented in Supplementary Figure S1, suggest that the control layout has no major impact on the performance of the multiplicative PMP method. The size of the plate remains a more important factor in this context.

3.2 Analysis of the RNAi HIV HTS assay

We applied the introduced mPMP algorithm to correct the RNAi HIV data generated during a genome-wide siRNA screen, which was aimed at studying HIV-host interactions. This screen was used by Carralot et al. (2012) to validate their diffusion correction model. To identify host factors involved in the interactions with HIV, an RNAi screening of human cells infected by HIV-1 and transfected with a genome-spanning siRNA library was carried out (for more details, see Supplementary Material S3 in Carralot et al., 2012). Carralot et al. showed that this screen was affected by multiplicative spatial bias, which was evident as edge effects. The original screen consisted of 68 plates with of size (16 x 24). Because columns 1 to 4 of all plates contained only non-target elements, their measurements were excluded from our analysis. Thus, our experiments were carried out using the 320 remaining measurements of each plate. The entire tested dataset is available at the following URL address: www.info2.uqam.ca/~makarenkov_v/HTS/downloads/RNAi_HIV_68.zip and the hit counts per well location represented in Figure 6a–f are available in Supplementary Tables ST1–ST6.

As the RNAi HIV screen is an inhibition assay, the hits correspond to low values of measurements. Consider Plate 7 of this assay (see Fig. 1b or Fig. 6g). First, we applied the Mann–Whitney U test to identify biased rows and columns present in this plate. The presence of spatial bias was detected in rows A to C, E and M to P as well as in columns 23 and 24 of Plate 7. The Kolmogorov-Smirnov test’s P-value for the additive PMP (aPMP) model was 0.0025, while the P-value for the multiplicative PMP (mPMP) model was 0.2196. The null hypothesis, $H_0$, here is that both unbiased raw measurements and corrected measurements come from the same distribution. Thus, the aPMP method provided strong evidence against the null hypothesis, while the mPMP method did not, at the selected significance level $\alpha$ ($\alpha$ was equal to 0.05 in our study). This result suggests that spatial bias in Plate 7 follows a multiplicative model.

Using the RNAi HIV experimental data, we compared the performances of seven data correction methods in terms of the number of hits (Table 1) and the data homogeneity, studied within Plate 7 (Fig. 6g–l) and within the overall hit distribution surface representing the number of hit counts per well location (Fig. 6a–f and Table 2). The seven compared methods were as follows: No Correction,
diffusion model removing multiplicative plate-specific bias (Carralot et al., 2012), conventional (additive) B-score (Brideau et al., 2003), assay-wise correction by Well Correction (Makarenkov et al., 2007; this procedure removes both additive and plate-specific biases across a given well location), plate-wise correction by multiplicative PMP, plate-wise correction by multiplicative PMP followed by assay-wise correction using Z-score normalization, and assay-wise correction by Z-score followed by plate-wise correction by multiplicative PMP. Our computations were carried out for four different hit selection thresholds consisting of 1%, 2%, 3% and 4.13% of hits (the last threshold was selected following Carralot et al., 2012). Table 1 shows that the conventional B-score correction drastically overestimates the number of detected hits compared to raw data. In contrast, the diffusion model, plate-wise correction by mPMP and the combined plate and assay-wise correction present hit totals close to that of raw data (see Table 1 and Fig. 6).

The application of the diffusion model to the RNAi HIV data led to a partial correction of the multiplicative edge effect affecting both Plate 7 and the assay’s hit distribution surface (Fig. 6h and b). However, this correction was not enough to pass the $\chi^2$ goodness of fit test for three of the four selected hit selection thresholds (Table 2). This test can be used in HTS to assess the deviation of the hit distribution surface from the expected (i.e. plane) surface (Makarenkov et al., 2007). The additive B-score technique removed an important part of the original edge effect at the expense of a significant increase in the number of detected hits (Table 1) and an inverse edge effect due to overfitting which can be observed within the corrected hit distribution surface (Fig. 6c). Even though the B-score method was able to remove spatial bias from rows C and O and column 23 of Plate 7 (Fig. 6i), it was by far the worst method in terms of the $\chi^2$ goodness of fit test used to assess the homogeneity of the hit surface (Table 2). Assay-specific bias correction improved the uniformity of the hit count surface by removing from it the patterns of edge effect. The corrected hit surface passed the $\chi^2$ goodness of fit test for all four hit selection thresholds.

However, an inverse edge effect pattern, similar to that introduced by B-score, can be observed on the corrected hit distribution surface (Fig. 6d). Assay-wise correction was also unable to correct a strong edge effect present in row A of Plate 7; this edge effect was apparently much more significant within Plate 7 than within the rest of the plates of this assay. Plate-specific bias correction via multiplicative PMP better corrected Plate 7’s edge effects (Fig. 6k), but still conserved the outlines of the original edge effect in rows O and P (Fig. 6e). All hit count surfaces computed after the plate-wise correction were also successful in passing the $\chi^2$ goodness of fit test (Table 2).

As both plate and assay-wise corrections passed the $\chi^2$ goodness of fit test for all four hit selection thresholds, they can be combined to obtain a more powerful bias correction technique. However, the order of their application is important as the results in Tables 1 and 2 suggest. While the results regarding the homogeneity of the hit distribution surface, reported in Table 2, give no advantage to one of the two methods, those regarding the total number of hits, reported in Table 1, show that plate-wise correction should precede assay-wise correction in the case of the RNAi HIV data. In fact, if the assay-wise correction precedes the plate-wise correction, a clear overestimation of the number of detected hits can be observed.

It is worth noting that the most suitable order of application of the plate and assay-wise corrections depends on the data only. For example, for a 5-plate assay presented in Supplementary Tables ST7–ST11 (see Supplementary Materials), the assay-wise correction

---

**Fig. 6.** Hit distribution surfaces and Plate 7 hitmaps for the following types of RNAi HIV data: (a,g) raw data, (b,h) data corrected by the diffusion model, (c,i) data corrected by additive B-score, (d,j) data corrected only assay-wise, (e,k) data corrected only plate-wise using mPMP and (f,l) data corrected both plate and assay-wise. The results for the hit selection threshold of $m_{1.219r}$ are depicted. Figure 1 description applies here
Table 1. Number of hits found in the RNAi HIV assay using: raw data (No Correction), data corrected by diffusion model, data corrected by additive B-score, data corrected only assay-wise, data corrected only plate-wise by mPMP, data corrected assay-wise and then plate-wise by mPMP, and data corrected plate-wise by mPMP and then assay-wise

<table>
<thead>
<tr>
<th>Method</th>
<th>No correction (raw data)</th>
<th>Diffusion model</th>
<th>B-score</th>
<th>Assay-wise correction</th>
<th>Plate-wise correction (mPMP)</th>
<th>Assay-wise plate-wise correction</th>
<th>Plate-wise assay-wise correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits</td>
<td>671</td>
<td>677</td>
<td>1271</td>
<td>851</td>
<td>708</td>
<td>944</td>
<td>750</td>
</tr>
<tr>
<td>Percentage</td>
<td>1036</td>
<td>1058</td>
<td>1595</td>
<td>1215</td>
<td>1050</td>
<td>1341</td>
<td>1079</td>
</tr>
<tr>
<td>Corrected</td>
<td>1239</td>
<td>1270</td>
<td>1719</td>
<td>1399</td>
<td>1247</td>
<td>1517</td>
<td>1284</td>
</tr>
</tbody>
</table>

Note: The selected thresholds, μ-1.348σ, μ-1.293σ, μ-1.255σ and μ-1.219σ, correspond to 1%, 2%, 3% and 4.13% of hits, respectively.

Table 2. χ² goodness-of-fit statistic (given for α = 0.01) for the hit distribution surfaces of the RNAi HIV assay computed after the application of the following data correction methods: No Correction, diffusion model, additive B-score, assay-wise correction only, plate-wise correction only by mPMP, assay-wise correction followed by plate-wise correction by mPMP, and plate-wise correction by mPMP followed by assay-wise correction

<table>
<thead>
<tr>
<th>Method</th>
<th>Critical value</th>
<th>μ-1.348σ</th>
<th>μ-1.293σ</th>
<th>μ-1.255σ</th>
<th>μ-1.219σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>No correction (raw data)</td>
<td>380.68</td>
<td>380.68</td>
<td>380.68</td>
<td>380.68</td>
<td></td>
</tr>
<tr>
<td>Diffusion model</td>
<td>366.26</td>
<td>408.91</td>
<td>477.51</td>
<td>529.91</td>
<td></td>
</tr>
<tr>
<td>B-score</td>
<td>369.97</td>
<td>421.64</td>
<td>480.30</td>
<td>537.12</td>
<td></td>
</tr>
<tr>
<td>Assay-wise correction</td>
<td>670.90</td>
<td>644.99</td>
<td>629.35</td>
<td>639.77</td>
<td></td>
</tr>
<tr>
<td>Plate-wise correction (mPMP)</td>
<td>338.38</td>
<td>299.31</td>
<td>279.12</td>
<td>268.71</td>
<td></td>
</tr>
<tr>
<td>Assay-wise plate-wise correction</td>
<td>315.28</td>
<td>329.45</td>
<td>345.81</td>
<td>368.40</td>
<td></td>
</tr>
<tr>
<td>Plate-wise assay-wise correction</td>
<td>300.07</td>
<td>282.85</td>
<td>280.95</td>
<td>268.21</td>
<td></td>
</tr>
</tbody>
</table>

Note: The selected thresholds, μ-1.348σ, μ-1.293σ, μ-1.255σ and μ-1.219σ, correspond to 1%, 2%, 3% and 4.13% of hits, respectively. Critical values are given in italic.

(Supplementary Figs SF2–SF6) should precede the plate-wise correction (Supplementary Figs SF7–SF11) because otherwise the assay-specific bias present in well locations (A,12), (B,12), (C,12), (D,12) and (E,12) cannot be effectively recognized (see Supplementary Tables ST7–ST11). Further validation of selected hits should be conducted through the structure-activity relationships (SAR) analysis and the subsequent clinical trials.

3.3 Analysis of ChemBank data

We first examined 100 experimental assays from ChemBank (Seiler et al., 2008), which is the most complete public small-molecule assay database, in order to determine the dominant type of plate-specific spatial bias affecting the four available HTS screening categories, including HTS-homogeneous, HTS-microorganism, HTS-cell-based and HTS-gene-expression assays (25 assays per screening category were analyzed (see Supplementary Table ST12 for the ChemBank IDs of the assays). The control wells were removed from all screens prior to bias detection. The proportion of assays per screening category, affected by additive bias, by multiplicative bias, and by an undetermined type of bias is reported. No assays containing no biased plates at all were found in this experiment.

![Fig. 7. Plate-specific bias detected across the four HTS screening categories available in ChemBank. 100 HTS assays (25 per screening category) were analyzed (see Supplementary Table ST12 for the ChemBank IDs of the assays). The control wells were removed from all screens prior to bias detection. The proportion of assays per screening category, affected by additive bias, by multiplicative bias, and by an undetermined type of bias is reported. No assays containing no biased plates at all were found in this experiment.](https://academic.oup.com/bioinformatics/article-abstract/33/20/3258/3868477)
cautiously since the application of error correction techniques on error-free data can introduce an additional bias that negatively affects the hit selection process (see for example the results of the multiplicative B-score method on error-free data in Figs 2–5). Thus, the application of spatial bias correction methods should be supported by statistical tests. Finally, we showed that the discussed methods for removing multiplicative spatial bias and the introduced general data correction protocol are effective in detecting and cleaning experimental data generated by screening technologies. For example, after analyzing the ChemBank data, we were able to determine that the additive type of spatial bias is dominant in homogeneous and microorganism HTS screens, while cell-based and gene-expression HTS assays are mostly affected by multiplicative spatial bias. Clearly, the screening category has a direct impact on the nature of spatial bias (additive vs. multiplicative; see Figs 7 and 8). In the future, it would be interesting to conduct some additional experiments in order to establish whether the type of spatial bias also depends on some technical and environmental factors which can affect experimental screening campaigns, such as reader and pipette malfunctioning, unintended variations in compound concentration associated with agent evaporation, or temperature, lighting and air flow fluctuations.

Mpendi et al. (2015) showed the importance of QC metrics (e.g. Z'-factor, Zhang et al., 1999) and per-plate data visualization for identifying systematic errors in experimental HTS, especially for data with a high hit rate. It would be interesting to compare in the future the performances of Z'-factor and the \( \chi^2 \) goodness of fit test, which was used in this paper to assess the deviation of the hit distribution surface from a plane surface. The advantage of the \( \chi^2 \) goodness of fit test is that it can be carried out when the control measurement information is unavailable for data at hand (as in the case of the RNAi HIV assay analyzed in Section 3.2), but Z'-factor can provide a better indication of the presence of spatial bias for secondary screens involving high hit rates.

The presented methods and data correction protocol have been implemented in the AssayCorrector package, which is freely available on CRAN.

4 Discussion

In this paper, we described three novel methods, called Non-Linear Multiplicative Bias Elimination (NLMBE), multiplicative Partial Median Polish (mPMP) and multiplicative B-score, for removing multiplicative spatial bias from experimental screening data. The performances of the new methods were assessed in simulations. These simulations confirmed that both NLMBE and mPMP outperformed the multiplicative B-score technique, which was prone to generating false positive hits. The NLMBE method yielded slightly better performances than mPMP in terms of the true positive rate, while mPMP was better than NLMBE in terms of the combined false positive and false negative rate. Taking into account that mPMP converges much faster than NLMBE, the former is recommended for correcting multiplicative spatial bias in HTS assays. The proposed NLMBE and mPMP methods correct only the measurements of rows and columns of a given plate in which spatial bias was detected by the Mann-Whitney \( U \) test. This is the main advantage of these methods, compared to B-score (additive or multiplicative) and other data correction techniques that modify all the measurements of a given plate even though spatial bias is present in only a few of them. This property of the new methods allows us to address efficiently the over-fitting issue.

Moreover, we presented a general bias correction protocol, which can be used by HTS researchers to remove both assay and plate-specific spatial biases. The plate-specific part of this protocol includes a new algorithm, based on the use of the additive (Dragiev et al., 2012) and multiplicative PMP methods as well as of the Kolmogorov-Smirnov two-sample test to identify the most appropriate (i.e. additive or multiplicative) spatial bias model for a given plate. We also propose to carry out the Mann–Whitney \( U \) test to detect the presence of both assay and plate-specific spatial biases. Importantly, the presented bias correction methods should be used cautiously since the application of error correction techniques on error-free data can introduce an additional bias that negatively affects the hit selection process.

Fig. 8. Plate-specific bias detected across 7 816 plates, including 3 001 344 gene expression profiles, from the Normalized L1000 mRNA profiling assay. The control wells were removed from all screens prior to bias detection. The proportion of plates affected by additive bias, by multiplicative bias, by an undetermined type of bias, as well as of those having no spatial bias at all, is reported. When the Mann-Whitney \( U \) test detected no any biased row or column in a given plate, the plate was reported as containing no spatial bias compared to 96.36% biased microplates found by Lachmann et al. using an algorithm which combines spatial autocorrelation detection and principal component analysis. Precisely, we established that 25.67% of the assay plates were affected by additive bias, 44.65% by multiplicative bias, 18.69% by an undetermined type of bias and 10.99% contained no bias (Fig. 8).

4 Discussion

In this paper, we described three novel methods, called Non-Linear Multiplicative Bias Elimination (NLMBE), multiplicative Partial Median Polish (mPMP) and multiplicative B-score, for removing multiplicative spatial bias from experimental screening data. The performances of the new methods were assessed in simulations. These simulations confirmed that both NLMBE and mPMP outperformed the multiplicative B-score technique, which was prone to generating false positive hits. The NLMBE method yielded slightly better performances than mPMP in terms of the true positive rate, while mPMP was better than NLMBE in terms of the combined false positive and false negative rate. Taking into account that mPMP converges much faster than NLMBE, the former is recommended for correcting multiplicative spatial bias in HTS assays. The proposed NLMBE and mPMP methods correct only the measurements of rows and columns of a given plate in which spatial bias was detected by the Mann-Whitney \( U \) test. This is the main advantage of these methods, compared to B-score (additive or multiplicative) and other data correction techniques that modify all the measurements of a given plate even though spatial bias is present in only a few of them. This property of the new methods allows us to address efficiently the over-fitting issue.

Moreover, we presented a general bias correction protocol, which can be used by HTS researchers to remove both assay and plate-specific spatial biases. The plate-specific part of this protocol includes a new algorithm, based on the use of the additive (Dragiev et al., 2012) and multiplicative PMP methods as well as of the Kolmogorov-Smirnov two-sample test to identify the most appropriate (i.e. additive or multiplicative) spatial bias model for a given plate. We also propose to carry out the Mann–Whitney \( U \) test to detect the presence of both assay and plate-specific spatial biases. Importantly, the presented bias correction methods should be used cautiously since the application of error correction techniques on error-free data can introduce an additional bias that negatively affects the hit selection process (see for example the results of the multiplicative B-score method on error-free data in Figs 2–5). Thus, the application of spatial bias correction methods should be supported by statistical tests. Finally, we showed that the discussed methods for removing multiplicative spatial bias and the introduced general data correction protocol are effective in detecting and cleaning experimental data generated by screening technologies. For example, after analyzing the ChemBank data, we were able to determine that the additive type of spatial bias is dominant in homogeneous and microorganism HTS screens, while cell-based and gene-expression HTS assays are mostly affected by multiplicative spatial bias. Clearly, the screening category has a direct impact on the nature of spatial bias (additive vs. multiplicative; see Figs 7 and 8). In the future, it would be interesting to conduct some additional experiments in order to establish whether the type of spatial bias also depends on some technical and environmental factors which can affect experimental screening campaigns, such as reader and pipette malfunctioning, unintended variations in compound concentration associated with agent evaporation, or temperature, lighting and air flow fluctuations.

Mpendi et al. (2015) showed the importance of QC metrics (e.g. Z'-factor, Zhang et al., 1999) and per-plate data visualization for identifying systematic errors in experimental HTS, especially for data with a high hit rate. It would be interesting to compare in the future the performances of Z'-factor and the \( \chi^2 \) goodness of fit test, which was used in this paper to assess the deviation of the hit distribution surface from a plane surface. The advantage of the \( \chi^2 \) goodness of fit test is that it can be carried out when the control measurement information is unavailable for data at hand (as in the case of the RNAi HIV assay analyzed in Section 3.2), but Z'-factor can provide a better indication of the presence of spatial bias for secondary screens involving high hit rates.

The presented methods and data correction protocol have been implemented in the AssayCorrector package, which is freely available on CRAN.

Funding

This work was supported by Le Fonds Québécois de la Recherche sur la Nature et les Technologies (grant no. 173878) and Natural Sciences and Engineering Research Council of Canada (grant no. 249644). We also thank three anonymous referees for their helpful comments.

Conflict of Interest: none declared.

References


Detecting and removing multiplicative bias in HTS


