JOURNAL OF APPLIED STATISTICS
2024, VOL. 51, NO. 6, 1210–1226
https://doi.org/10.1080/02664763.2023.2248413

Taylor & Francis
Taylor & Francis Group

APPLICATION NOTE

Classifying contaminated cell cultures using time series features

Laura L. Tupper[a], Charles R. Keese[b] and David S. Matteson[c]

[a]Mount Holyoke College, South Hadley, MA, USA; [b]Applied BioPhysics Inc., Troy, NY, USA; [c]Cornell University, Ithaca, NY, USA

ABSTRACT

We examine the use of time series data, derived from Electric Cell-substrate Impedance Sensing (ECIS), to differentiate between standard mammalian cell cultures and those infected with a mycoplasma organism. With the goal of easy visualization and interpretation, we perform low-dimensional feature-based classification, extracting application-relevant features from the ECIS time courses. We can achieve very high classification accuracy using only two features, which depend on the cell line under examination. Initial results also show the existence of experimental variation between plates and suggest types of features that may prove more robust to such variation. Our paper is the first to perform a broad examination of ECIS time course features in the context of detecting contamination; to combine different types of features to achieve classification accuracy while preserving interpretability; and to describe and suggest possibilities for ameliorating plate-to-plate variation.

ARTICLE HISTORY
Received 24 February 2022
Accepted 30 July 2023

KEYWORDS
Time series classification; feature-based classification; contamination of cell cultures; biophysics; electric cell-substrate impedance sensing

1. Introduction

1.1. Data challenge: cell contamination

The study of cells in culture is a vital component of biological research, revealing cells' physical morphology, their patterns of growth and progression through the life cycle, and their responses to stimuli and the environment. Yet there is a reproducibility crisis in cell culture research, in large part due to misidentification and contamination of cell samples (see [14]). One common issue is the contamination of mammalian cells with other microorganisms such as *Mycoplasma*, which can flourish in the medium used to grow the cells and create misleading results.

In this paper, we address the issue of contaminated cell cultures as a classification problem, with two equally important goals: to achieve high classification accuracy of cells as infected or uninfected, but also to produce results that are easy to interpret, visualize, and intuit in the context of the application. We draw on an automated, non-invasive data collection method, *electric cell-substrate impedance sensing*, to generate time series corresponding

CONTACT Laura L. Tupper ltupper@mtholyoke.edu Mount Holyoke College, 50 College St., South Hadley, MA 01075, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02664763.2023.2248413.

to individual cell cultures. Several typical methods for classifying high-dimensional data – such as multivariate logistic regression, PCA, and any classification process based on a large number of features – are not appropriate since they are in conflict with the goal of easy interpretability and visualization. Instead, we constrain our classification procedure to only two dimensions, drawing on feature-based methods that use characteristic features derived from these time series. We find that it is indeed possible to fulfill both of our goals, obtaining high classification accuracy with relatively simple, low-dimensional methods.

The structure of the paper is as follows. In the current section, we introduce the data collection method and the specific datasets used in this paper. Section 2 describes our approach: the general methodology of low-dimensional feature-based time series classification, and the types of features we generate from the data, with references to the previous studies that suggested these features. In Section 3 we classify individual cultures as infected or uninfected and examine the performance of different types of features, while Subsection 3.4 extends the investigation to different cell types and addresses the possibility of experimental variation. Conclusions and suggestions for future research appear in Section 4. The online supplementary materials provide more details on the methodology of data collection, a complete list of features investigated, and additional classification performance results.

1.2. Relevant data

1.2.1. Electric cell-substrate impedance sensing

Electric cell-substrate impedance sensing, or ECIS®, is an established tool for measuring the behavior and characteristics of cells over time. An introduction can be found in [11]; since then, the technique has been used in many studies in a variety of biological contexts. Lukic and Wegener [18] contain a survey of many of these studies, while [13] provide an overview of the use of ECIS in cancer studies.

In parallel to this, there has been considerable scholarship on using cell *images* for phenotype classification; for example, [21] emphasize the importance of dimension reduction in this context, while [8] extracts features based on the cell images' texture. ECIS serves as a useful alternative to optical or image-based assessment of cell cultures, as it requires no chemical labels or markers to be applied to the cells and no image preparation process, but can identify the same types of changes; for example, [7] demonstrate that ECIS-based and optical determinations of cell confluence coincide. Thus we investigate similar questions of dimension reduction and feature-based classification while focusing on ECIS-specific features.

The outline of the ECIS technique is as follows. First, an experimenter inoculates a plate containing several wells with a cell culture. In the bottom of each well are electrodes, smaller than the entire surface of the well. The plate is placed in a machine that passes a weak AC current through the electrodes, at a chosen frequency, and measures the impedance. With appropriate equipment, the impedance can also be decomposed into its two complex components, resistance and capacitance. As the cells multiply and move, they cover more of the electrodes, and the nature of their attachment to the medium underneath them may also change. Eventually they reach confluence, when there is complete coverage of cells across the entire well. All of these processes change the observed impedance and its components at various frequencies.

The ECIS process is non-invasive, and does not damage the cells or change their morphology, allowing for sustained observation of the same culture. If desired, however, the machine can *wound* the culture, by passing a high-voltage current through the electrodes. This process kills most or all of the cells located on top of the electrodes. We can then observe further changes in impedance as the dead cells are replaced by living cells moving in from outside the edges of the electrodes. ECIS observations can be considered as time series; these time series are high-dimensional if we consider each well to have measurements on multiple frequencies linked by time point, or if we consider both components of impedance at each time point.

Many studies have used ECIS data to distinguish between specified cell types, or to observe differences in a cell before and after it undergoes some process, such as electroporation [27], damage [12], transformation into cancer cells [22], or progression to a new phase of the cell cycle [31]. But quantitative discussion of the data has been sparse. Gelsinger et al. [10] appear to be the first to examine quantitative features of ECIS data across multiple cell lines, with the goal of identifying the cell types in unknown or potentially-mislabeled cultures. Zhang et al. [33] examine contamination data as we do in this paper, comparing long-memory behavior and the timing of the confluence stage in infected and uninfected cells. We build on this work by bringing in many other types of features, with a quantitative comparison of their efficacy in identifying contamination in various cell types; this paper expands on our conference presentation in [29].

1.2.2. The current dataset

In this paper, we use data provided by Applied BioPhysics, Inc., generated using the ECIS $Z\theta$ machine. Here we describe the details of the datasets.

Cells are grown on arrays, or plates; each plate contains 96 wells, allowing the observation of multiple cell types and growth conditions within a single plate. At each time point, the ECIS $Z\theta$ equipment generates four measurements: Z, impedance; R, the resistive component; C, the capacitive component; and H, the time in hours corresponding to the current time index. These measurements are obtained for each well using several different AC frequencies, from 500 to 32000 Hz.

Note that these wells are fabricated with two electrodes which both supply the current and measure the potential, so that the electrodes' own impedance is included in the measurement. This is a potential source of variation; but an examination of empty wells shows that this electrode impedance does not vary greatly across wells even from different plates. The electrode impedance is also essentially constant over time, so that many of the features described below will be unaffected.

The cultures in our dataset are of two types: the target cell line, and the same cell line contaminated with a mycoplasma organism. We first examine MDCK II cells, some of which are contaminated with *M. hominis*. Cells are grown on one of two substrates: either with a gelatin coating, or with an adsorbed layer of BSA (bovine serum albumin). Some wells of each type are left empty, containing substrate but no cells, to provide a baseline.

As an extension to our original investigation, we also examine cultures from a different cell line, BSC-1, some of which are infected with *M. hominis*, and MDCK II cells infected with a different species of mycoplasma, *M. hyorhinis*. A description of these observations can be found in Subsection 3.4.

2. Methodology: feature-based classification

2.1. The feature-based approach

In feature-based classification, instead of calculating similarity or distance 'pointwise' between individual time points in the time series, we reduce the effective dimension by generating a limited set of features from the original data. We may then perform classification or clustering using these features, or a subset of them. There is an extensive body of literature on feature generation and selection; an overview and typography of many approaches appears in [6]. Maharaj et al. [20] provide a discussion of feature-based approaches specifically for the classification of time series data. One strength of the feature-based approach is the reduction of dimensionality, as argued in [30], which can avoid many of the problems involved in high-dimensional clustering (for examples, see [1]). We also avoid the problematic requirement that all the time series under consideration be exactly the same length, as mentioned in [15].

The feature-based approach is designed to allow for visualization and interpretation, since the user need only consider a few features to understand why an observation is classified in a particular way. In particular, we restrict the problem to two dimensions by using *pairs* of features to perform classification. The purpose of this constraint is to allow visualization of the classification regions in a two-dimensional feature space, along with the distribution of feature values across observations.

In this paper, we generate a variety of features from the ECIS time courses. We have attempted to designate features that reflect characteristics of ECIS data mentioned as useful in previous studies – though these previous studies did not attempt to combine different types of features. We have also included some features based on time series analysis and those aimed at capturing particular stages of the cells' behavior. All computation was performed using R [24] with `tidyverse` packages [32]; estimation details for individual features are given in the supplementary materials.

2.2. Features derived from ECIS data

While the shape of the ECIS time series depends on the cell line, the substrate on which it is grown, and any contamination, there are general patterns that inform our choice of features. Immediately following the initial inoculation of the well, there is an attachment phase as the cells attach themselves to the substrate, then a growth period as cells reproduce and move to cover the electrode. These correspond to increasing resistance values in the data. Once the cells reach confluence – possibly following a drop in resistance from an initial peak – large-scale movement and spreading becomes impossible, and we see a long period of more consistent values. During this phase, the cells exhibit *micromotion*, small movements that cause fluctuations in the resistance, and may also show a gradual linear trend. The exact time at which the cells reach confluence, and the direction of the overall trend in resistance during confluence, depend on the cell type. Across cells, however, the period between 24 and 36 hours can be safely considered a 'confluent window'. After considerable time, usually beyond 72 hours, the cells may begin to senesce, leading to more substantial changes in resistance.

The types of features we examine are listed below, along with references to the use of similar features in existing literature.

2.2.1. Level of resistance at a time point

Many analyses of ECIS data simply discuss the overall level of impedance (or one of its components) as contrasted between cell types. Accordingly, we include several features that reflect the level of resistance at a particular time, measured separately at each frequency.

2.2.2. Rate of change

Park et al. [22] compare cell lines with their transformed (cancerous) versions, and observe that resistance of the cancerous cells 'increased more rapidly' than that of their noncancerous counterparts. A similar feature appears in [16]: they discuss the 'initial increase in resistance' as distinguishing between control cells and cells that are treated with compounds encouraging attachment. We define two features to reflect this rate of increase of resistance early on. First, we simply record the resistance level at 2 hours after inoculation; this time is during the early attachment and spreading stage, and the resistance at this point gives a measurement of how quickly the cells are spreading. We then calculate the difference between the levels at 7 hours and 2 hours after inoculation, which effectively measures the slope during the growth period but after the very early attachment phase. It is also evident that the early growth in resistance is not linear. Accordingly, we fit a quadratic to the first 2 hours of values after inoculation, and extract the first- and second-degree coefficients from the fit. All these early measurements may be particularly useful in contexts where results must be obtained quickly, as in [25], where practitioners have only a few hours to determine whether stem cells will be suitable for a recipient before the cells become non-viable.

2.2.3. Peaks and confluence

The confluence level of resistance is used in [12,16], while a late-stage measurement of capacitance appears in [2]. We measure the resistance level at 24 hours after inoculation, which serves as an estimate of the level at confluence. Park et al. [22] also note that cancerous cells 'peaked higher' than noncancerous cells: we adopt this feature by calculating the maximum level of resistance over the first 24 hours, and over the first 48 hours. Using a uniform time window is important, since some types of cells have upward trends in resistance during confluence and may reach their maximum value at the end of the observed window. As an alternative that does *not* reflect such behavior, we also measure the time and value of the *first peak* following inoculation in a smoothed version of the time series.

2.2.4. Wounding and recovery

Several studies address cells' responses to wounding, or 'the ability of cells …to close an experimental lesion,' as [18] put it. In the same vein, Heijink et al. [12] observe resistance levels when confluence is regained after cells are damaged with cigarette smoke. Stolwijk et al. [27] provide a similar example of examining cells' recovery from an event, though in this case the event is non-fatal electroporation. This study notes variation in time to recovery 'dependent on cell type and age of the culture.' Time to recovery also appears in [12], as does the reduction in resistance at one to two hours after wounding. In our dataset, wounding is performed at approximately 48 hours after inoculation, and we define several features to reflect post-wounding behavior. First, we record the minimum level of resistance at (or just following) wounding, since the residual resistance from dead cells can provide information about their morphology. To represent the rate of recovery after wounding, we use the difference between the levels at 57 and 52 hours after inoculation. In addition, we

take the ratio of this post-wounding slope to the slope during the initial growth stage. At high frequencies, we observe a particular pattern in the post-wounding resistance levels: after the sharp drop in resistance in response to wounding, there is a period of swift growth to a local peak, followed by a drop to a more moderate level. Although the final resistance levels of infected and uninfected cells may be similar, the characteristics of this first post-wounding peak are different. Accordingly, we include features corresponding to both the value of this peak and the time when it occurs.

2.2.5. Post-confluence micromotion

Several early papers by Giaever and others, such as [11], examine cells' micromotion behavior, which can be observed in ECIS data as small fluctuations after a stable level has been reached. Lukic and Wegener [18] mention examining micromotion on a 30-minute window, which helps account for any lingering overall trends in level. There appears to be no consensus among researchers on the best way to quantify micromotion. We measure small-scale variation during confluence by looking at the variance and standard deviation of each time course during the confluent window (24 to 36 hours after inoculation), after differencing the values to remove trend. For comparison, we also include the estimated error variance after fitting classical ARIMA models, discussed below.

2.2.6. Time series models

We investigate classical ARIMA models, focusing on the stable confluent window between 24 and 36 hours. Using automated selection with AIC, we find the preferred ARIMA order over this range for each time series separately; this procedure almost always suggests first-order differencing, which agrees with our observation that the time series have a linear trend over this window. We convert each flexible-order ARIMA to an approximate AR(∞) representation, as seen in [23], and extract the first five coefficients. (With the exception of a few outliers, the coefficients beyond this point become too small to be useful in classification.) As a comparison, we also fit two simple fixed-order ARIMA models, $ARIMA(1, 1, 0)$ and $ARIMA(0, 1, 1)$, to each series and extract their fitted coefficients. The estimates of error variance from all of these models can also be useful in reflecting micromotion, as introduced by Tong and Dabas [28]. Using model coefficients and errors for clustering has appeared in a variety of time series application areas (for example, [4,19]) and is discussed more fully in [20].

2.2.7. Autocorrelation measures

To analyze the time series' short-term memory without fully fitting a model, we find the values of the ACF and PACF at small specific lags for the differenced series, as well as the Box-Pierce and Ljung-Box statistics, which measure the overall amount of autocorrelation in a time series. An ACF-based distance measure for time series clustering is described fully in [9], and related examples and extensions can be found in [3,30] (introducing the PACF as well as the ACF), and [5] (extending the method to fuzzy clustering). Though these methods have been used in other time series areas, they appear to be new to the specific context of ECIS.

2.2.8. Long-memory behavior

Multiple papers have noted potential long-memory behavior in ECIS time series. The DFA (detrended fluctuation analysis) approach to quantifying micromotion appears in [17] in examining cells' response to a toxin, and in [26] when comparing cells before and after a phenotypic transition. Recently, Zhang et al. [33] have performed a more involved analysis of the regime change between early growth and confluence. Their *Growth-to-Confluence Detector* assumes that early growth follows a nonlinear growth curve with heteroscedastic noise, while the confluence stage follows a long memory model. Following their methodology, we extract the estimated long memory parameter from post-confluence data, as well as the estimated time of the change point from growth to confluence behavior.

2.3. Variations of feature types

We calculate each of the features above on the time course of resistance. Basic feature types (maxima, values at a set time, and features combining these) are found both for the raw time courses and for a smoothed version, using a rolling windowed average over 5 consecutive time points. This windowing is intended to make the feature values robust to momentary noise in the data. For robustness to changes in laboratory conditions over time, we also calculated *normalized* time series relative to the behavior of empty wells on the same plate: details of this process can be found in the supplementary materials. After calculating many features both on the original and normalized time courses, we have found that this normalization has only minimal impact; thus more advanced features are calculated only using the original data.

All features are calculated for all frequencies available in the dataset, since measurements at different frequencies may reflect different components of cell behavior. This idea is discussed in [18], and some studies make reference to particular frequencies of interest; for example, [31] find 60 kHz to be most useful in their study of the cell cycle.

2.4. Classification process

Just as we restrict the dimension of the problem for interpretability, we also require that the classification process yield simple, easy-to-visualize decision regions in the feature space, like those in Figure 2. This requirement (along with the relatively small training dataset) precludes many advanced learning methods. Instead, we use three classical methods: classification trees, linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA). A thorough discussion of these methods appears in [10], including an exploration of different parameters for discriminant analysis – which are found not to have a substantial impact on the results. In our own analysis, we observe that LDA generally performs the best of the three methods, or close to it; and in general LDA may be preferred for these fairly small samples since there are fewer parameters to estimate. Accordingly we present detailed results based on LDA. Further notes on classification methods can be found in the supplemental materials.

We perform classification using each possible pair of features, in turn; from the 546 features and feature variations, we obtain over 148,000 possible feature pairs. Given a feature pair, we find the classification accuracy as follows: first, we divide the observations randomly into training and testing sets, each containing some infected and some uninfected

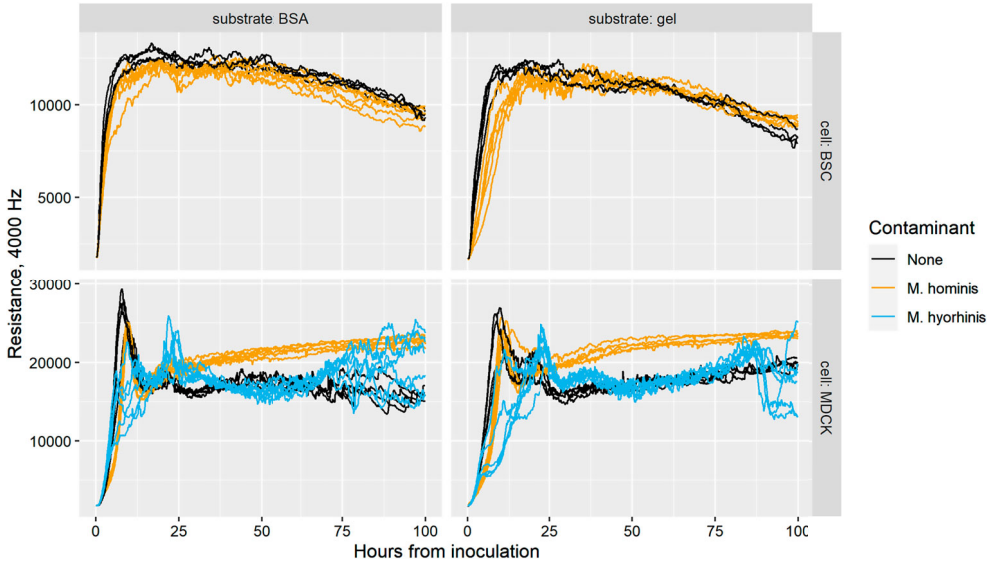

Figure 1. Sample ECIS data: time courses of resistance for different cell types.

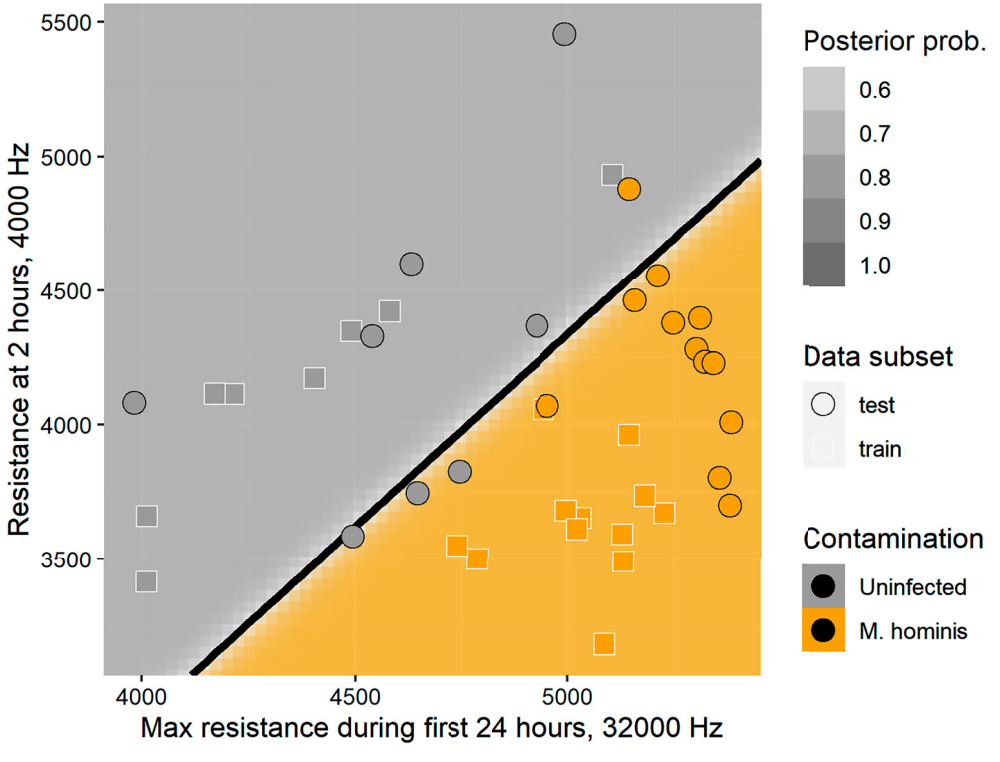

Figure 2. LDA classification regions based on one feature pair, with perfect classification of the test set.

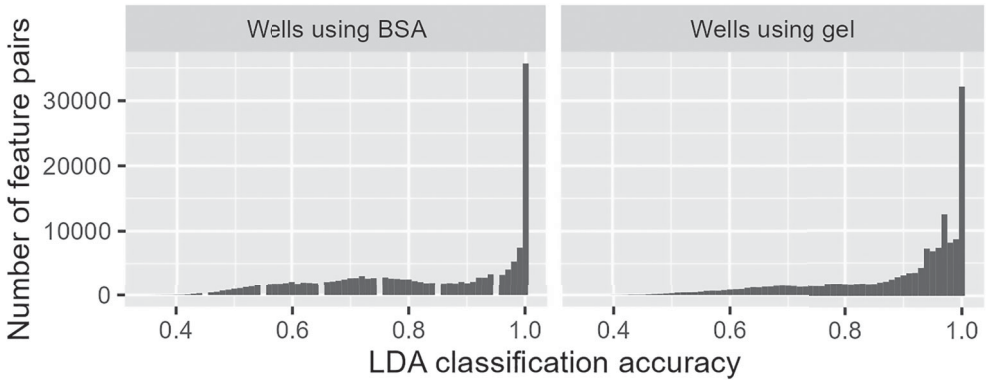

Figure 3. Histogram of classification accuracy values for each pair of features: many yield perfect accuracy.

wells. We then train an LDA classifier based on the training observations' feature values, use it to classify the testing observations as infected or uninfected, and record the success rate of these test-set classifications. We then repeat the process with a different training/testing split, a total of 10 times. The performance values discussed below use the average test-set accuracy over all 10 splits.

3. Application and results

3.1. General classification results

An example of the LDA classification results is shown in Figure 2, using two simple features: resistance at 2 hours after inoculation (measured at a frequency of 4000 Hz) and the maximum resistance during the first 24 hours after inoculation (at 32000 Hz). The classifier uses the training observations (square points) to generate a posterior probability of contamination for each point in the feature space, and a linear division of the feature space into infected and uninfected regions. The test observations (round points) are all correctly classified based on their coordinates in this feature space, and we can see that there is very clear separation between contaminated and uncontaminated cultures in terms of these feature values.

In general, classification accuracy is high. There is no 'best pair' of features: many different feature pairs offer strong, and in many cases equal, performance. On BSA-treated wells, about 24% of all feature pairs yield perfect classification accuracy. The proportion of perfect-accuracy pairs is about 21% on gel-treated wells, though there are more feature pairs that yield *near*-perfect accuracy on gel, as shown in Figure 3.

The protein coat on which the cells are grown does appear to affect their behavior and the differences between infected and uninfected cells: some features are more effective for classification of cells grown with BSA, while others are more effective on gel-treated wells. Since it is a simple matter in practice to grow cells using one or the other substrate, high performance in either case is equally useful; when giving performance values, we will indicate on which substrate the performance is achieved.

This is a promising indication of the effectiveness of ECIS measurements for distinguishing contaminated cultures, using simple features in a low-dimensional space that is easy to visualize. It is worth noting, though, that the sample size is small, and the problem becomes more difficult once we introduce the possibility of experimental variation across plates, as discussed in Subsection 3.4.

3.2. Feature combinations

Figure 4 shows a heatmap of performance values for a subset of feature pairs, with the brighter tiles representing feature pairs with higher accuracy. This subset illustrates the high performance that can be achieved with many different features, but also suggests rules for combining features, based on the 'sub-diagonals' of poorer performance that can be seen throughout. Some of these low-performance areas correspond to pairs consisting of a feature and a modified version of the same feature: for example, *at2hR32000* (resistance at 2 hours, 32000 Hz) paired with *w5c2hR32000* (resistance averaged over a window at 2 hours, 32000 Hz). These feature modifications do not greatly change the information in the feature, so that pairing a feature and its modified version is essentially equivalent to using a single feature for classification. This behavior suggests that windowing is optional, though it may still be desirable for robustness against recording errors. Other low-performance cases appear when the pairing consists of the same feature calculated on two different versions of the data: the original time series and the version normalized to empty wells (features beginning with 'EN'). Again, this shows that normalizing the time series has only a minimal effect on the features calculated from it.

To a lesser extent, we also see relatively weak performance when a feature is paired with itself at a similar frequency, or with a modification of the same feature at a similar frequency. For example, *at2hR32000* and *at2hR16000* are not an effective pair, but each yields perfect classification accuracy if paired with *at2hR500*. This provides further evidence that more divergent frequencies provide more useful information, by reflecting different characteristics of the cells' behavior.

3.3. Individual feature performance

We can obtain a simple measure of the effectiveness of individual features by averaging the classification accuracy over all pairs involving a given feature. A few notable points are described below, and can be found in table format in the supplementary materials.

3.3.1. Early behavior

High classification accuracy can be achieved with features that use only the first several hours of data following inoculation. The fitted quadratic coefficients for the first two hours of the time series have excellent average accuracy for low frequencies (averaging above 0.99 on gel for frequencies up to 2000 Hz), while a simple difference between the resistance levels at 7 hours and 2 hours after inoculation is also effective (average accuracy 0.99 on gel at 2000 Hz). Even the simple level at two hours performs fairly well in this case (for example, 0.98 on gel at 500 Hz), though this early level can be sensitive to variation in equipment or environmental conditions at inoculation.

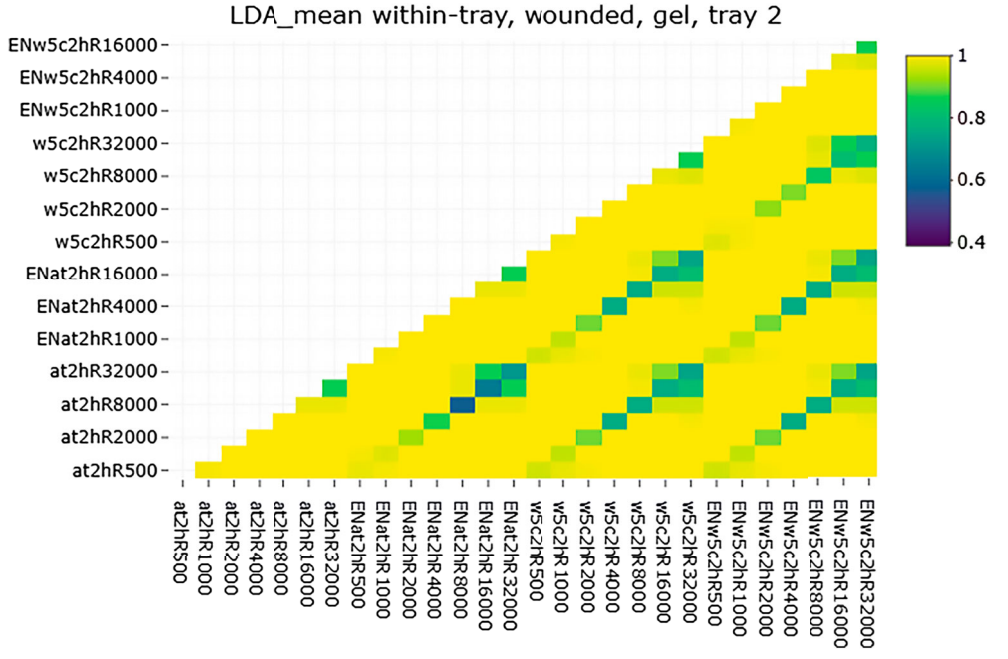

Figure 4. Heatmap of classification accuracy (on gel) for a few feature pairs, showing 'sub-diagonals' of feature pairs with poorer performance. The highest pair accuracy (bright yellow) is 1, while the sub-diagonal low-performance pairs (dark green/blue) have accuracy rates around 0.7, ranging as low as 0.57.

3.3.2. Peak values

The value of the post-wounding peak in resistance is very effective, with an average accuracy above 0.99 on gel for all frequencies up to 2000 Hz. The value of the initial peak after inoculation can also be useful (0.99 on gel at 2000 Hz). Identifying this first peak is slightly more effective than using the maximum value over a fixed time period (the best performance of this type is 0.98 on gel at 2000 Hz): in many cases these two features will identify the same peak, but for cell types with a post-inoculation peak and drop followed by an upward trend, using the maximum over a fixed window may inadvertently reflect the later post-confluence behavior rather than the initial peak.

3.3.3. Time series features

The coefficients from ARIMA models (fitted to the post-confluence period when the time series can be considered integrated of order 1) do not perform well in general. The cell cultures do display some short-memory behavior, and their fitted ARIMA models are substantially different from white noise and from models fitted to empty wells; but infected and uninfected cells do not have reliably distinct model coefficients. (Aside from the constant coefficients, which are arguably overly-complicated ways of describing the trend, the best performer is the $\hat{\theta}$ coefficient from a fixed ARIMA(0,1,1) model at 32000 Hz, with an average accuracy of 0.94 on gel. Among coefficients when the order of the ARIMA is not manually chosen, the best performer has an average accuracy of only 0.84.) Specific

values of the ACF and PACF seem to be a more effective way of characterizing the short-memory behavior; for example, the second lag of the PACF at 32000 Hz has an accuracy of 0.97 on gel. (Portmanteau test statistics are less useful, with a top performance of 0.88 on gel at 8000 Hz.) Notably, although ARIMA coefficients are not strong performers themselves, fitting ARIMA models can provide an alternative means of estimating local noise as a measurement of cell micromotion after confluence.

3.3.4. Post-confluence behavior

We measure cells' post-confluence behavior in two main ways: micromotion as reflected in the local variability of the time series, and long-memory behavior. Both seem to be an effective basis for classification. We obtain average accuracy of 0.97 or better on gel with any of several measures of local variability (the sample variance or standard deviation of the differenced series; or the estimated $\hat{\sigma}^2$ from an ARIMA model, whether fixed as order (0,1,1), fixed as order (1,1,0), or with the order chosen automatically for each series). Notably, all of these local variability measures perform best at the highest frequency, 32000 Hz. The estimated long-memory parameter, meanwhile, gives an accuracy of 0.98 on gel at 2000 Hz.

3.3.5. Post-wounding recovery

Features describing the cells' recovery from a wounding pulse can be very effective. For example, the level at 57 hours minus that at 52 hours represents the rate of recovery shortly after wounding; this has an accuracy of 0.99 on BSA at 500 Hz. It is also effective to compare this post-wounding recovery rate to the initial growth rate: the *ratio* of (57 – 52 hours) to (7 – 2 hours) has an average accuracy of 0.99 on gel for all frequencies up to 4000 Hz.

3.4. Extensions of the problem

3.4.1. Cell and contaminant types

We extend our analysis by examining cultures taken from a different cell line, as well as those infected with a different species of mycoplasma. For reasons of space, a detailed description is not included here; we focus on comparison with the original results, which used MDCK II cells with *M. hominis*. For a cell comparator, we use BSC-1 cells, some infected with *M. hominis*. We also examine MDCK II cells infected with another type of mycoplasma, *M. hyorhinis*. It is evident that the classification problem depends on both the type of cell and the species of mycoplasma; features that perform well for one cell-contaminant combination may not do so for others.

To distinguish MDCK II cells from those infected with *M. hyorhinis*, among the strongest-performing individual features are those focusing on the early peak behavior. In this case, however, it is more effective to use the maximum level in the first 24 or 48 hours after inoculation, rather than attempting to find the first peak as with *M. hominis*; this seems to be because the early growth period is noisier, leading to false local maxima before the real post-inoculation peak occurs. The other notably effective feature type here is the difference between the level at 57 and 52 hours after inoculation, a representation of the rate at which the cells recover from wounding. Since this also appeared as an effective feature for detecting *M. hominis* infection, wounding cells and monitoring their recovery may be a particularly useful approach when it is not certain what kinds of contaminants may be present.

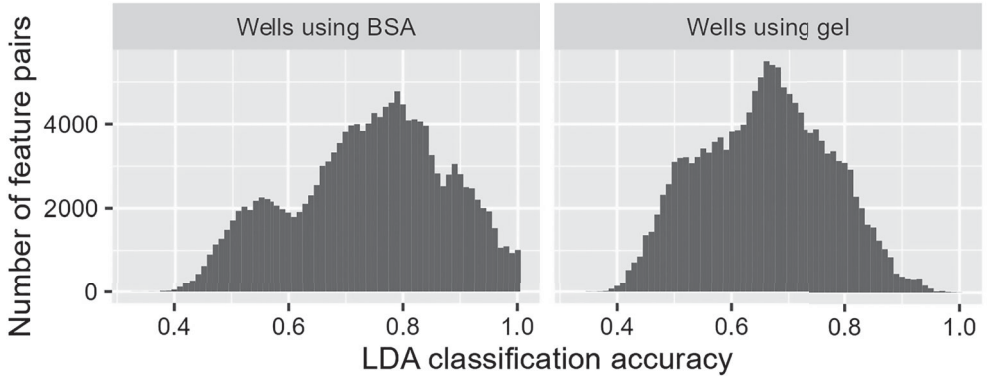

Figure 5. Histogram of classification accuracy values for each pair of features on BSC cells.

For the BSC-1 cells the problem is overall more difficult: as shown in Figure 5, far fewer feature pairs are able to give good classification performance, and classification appears to be easier for BSA-treated cultures than for those grown on gel. BSC-1 cells have a different growth and confluence pattern than MDCK II cells, notably without the large peak prior to confluence: as a result, the early levels and first-peak features that worked so well for MDCK II cells are not effective here. Instead, the features with high average performance for BSC-1 cells (above 0.91 on either protein coat) can be characterized as two types: coefficients from a quadratic fit to the first two hours of growth, or measurements of local variation/micromotion after confluence. These results underline the importance of establishing the most effective and characteristic features for individual cell lines, and of exploring a variety of possible feature types.

3.4.2. Experimental variation

In laboratory use, the cell cultures that require classification may often be grown on a separate plate than the training set of cultures known to be infected or uninfected. The test set may even be measured at a substantially different time, or under different conditions. It is therefore desirable to examine the robustness of classification to variation across plates, equipment variation, and experimental variation in general.

To this end, we examine four separate plates of MDCK II cells, each plate containing some uninfected and some infected cultures (with *M. hominis*) taken from the same frozen source, but grown and measured at a separate time. We can immediately see that plate-to-plate variation exists, and makes the classification problem substantially more difficult. Figure 6 shows the same set of features and classification procedure as Figure 2, but here, all observations on one plate are used as the training set, while the test observations come from a different plate. The LDA classifier is able to achieve perfect classification accuracy within a single plate, but cannot do so when working with two separate plates.

In the cross-plate classification problem, the accuracy rates are markedly lower than for the single-plate analyses above. Some results are consistent with those discussed above: ARIMA coefficients still do not perform well, and pairing a feature with its own modified version (at the same frequency) is ineffective.

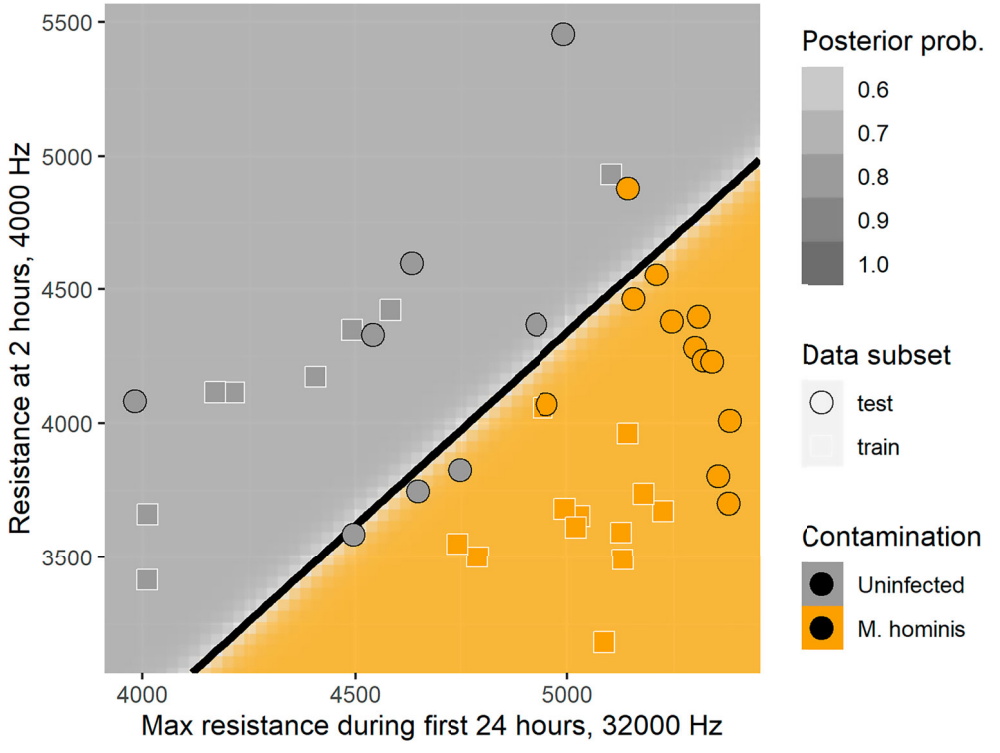

Figure 6. LDA classification regions, with training set from one plate and test set from a second plate.

Figure 7. Histogram of classification accuracy values for each pair of features, with training and test observations from different plates. Performance is generally much worse than for within-plate classification.

A noticeable standout is the '7 hours minus 2 hours' family of features, which reflect the rate of the initial cell growth after inoculation. Though simple, this feature type performs well when paired with several other types of features, or even with itself at two distinct frequencies; for example, it yields an average accuracy of 0.94 on gel-treated cultures at 4000 Hz. Such a feature draws on the information provided by the early growth

stage while accounting for the level reached during the first two hours, making the feature more robust to the initial conditions of each cell culture. This may provide greater robustness to experimental differences such as the initial number of cells at inoculation or the starting temperature of the cells. Features that involve differencing or comparison across multiple time points in a given well can also control for possible level differences between wells due to equipment variation.

In general, features focusing on later-stage behavior also appear to be more robust to plate-to-plate variation. The process of each culture reaching confluence, or being wounded with the same high-voltage current, appears to equalize the effects of initial conditions. Indeed, features measuring local variability after confluence are among the top overall performers (such as $\hat{\sigma}^2$ from various ARIMA models, with accuracy rates of about 0.92 on BSA at 500 Hz). These are promising indicators of useful features, though more data will be needed to characterize plate-to-plate variation more fully and determine which features are most robust to it.

4. Conclusion and discussion

This paper demonstrates a methodology for differentiating between standard cell cultures and those infected with a mycoplasma organism, by comparing features of their ECIS time courses. Detecting contamination in this way is particularly desirable since the ECIS data gathered during the process can also be used to investigate other scientific questions about cell morphology, without requiring a separate tool, or labeling or destroying cells.

We have determined that high classification accuracy can be achieved in combination with easy visualization and interpretation, by using a two-dimensional feature space and a straightforward classification algorithm such as LDA. In the basic problem, we achieve very high accuracy rates using any of a variety of feature pairs. Effective pairs usually combine features measured at different frequencies, which reflect different characteristics of the cells. Several different types of features can be of use, including those reflecting initial growth rates, peak resistance levels, post-confluence micromotion, recovery from wounding, and long-memory behavior. We have also seen, however, that the classification efficacy of specific features depends on the type of cell and the type of contaminant under investigation.

Based on our current dataset, we can also see that there is substantial variation across plates. We have already shown that certain features appear to be more robust to this 'plate effect' than others. Especially promising are features that reduce the effect of initial conditions: either by ignoring initial impedance levels, by waiting for the confluence stage to be achieved, or by wounding all the cultures and observing their recovery. Since experimental variation is to be expected in practice – where cultures will be grown using different equipment, at different times, and in multiple laboratories – this is an important topic for further investigation.

While few features show clear separation between all infected cells (across multiple plates) and all uninfected cells, we see that some features show separation within each plate and a consistent direction of difference across plates. If it is possible to normalize the data for each plate, such features would be effective for cross-plate classification. In a situation where several suspect sources of cells are being compared at once, this could be done by inoculating some wells on each plate with "baseline" cells from a single source

already known to be uninfected. But where experiments are being performed over a long time, or baseline cells are not locally available, some other methodology would be necessary to standardize the plates. In future, we intend to acquire a larger dataset in order to be able to study this plate effect more fully.

Disclosure statement

Charles R. Keese is Chief Scientific Officer of Applied BioPhysics, Inc., which manufactures the ECIS $Z\theta$ machine used to collect the data described in this paper.

Funding

The authors gratefully acknowledge financial support from the Cornell University Institute of Biotechnology, the New York State Foundation of Science, Technology and Innovation (NYSTAR), a Xerox PARC Faculty Research Award, National Science Foundation Awards 1455172, 1934985, 1940124, and 1940276, USAID, and Cornell University Atkinson Center for a Sustainable Future.

References

[1] C.C. Aggarwal, A. Hinneburg, and D.A. Keim, *On the Surprising Behavior of Distance Metrics in High Dimensional Space*, in *Database Theory – ICDT 2001. ICDT 2001. Lecture Notes in Computer Science*, J. Van den Bussche and V. Vianu, eds., vol 1973. Springer, Berlin, 2001, pp. 420–434. doi: 10.1007/3-540-44503-X_27

[2] P.O. Bagnaninchi and N. Drummond, *Real-time label-free monitoring of adipose-derived stem cell differentiation with electric cell-substrate impedance sensing*, Proc. Natl. Acad. Sci. 108 (2011), pp. 6462–6467.

[3] J. Caiado, N. Crato, and D. Peña, *A periodogram-based metric for time series classification*, Comput. Statist. Data Anal. 50 (2006), pp. 2668–2684.

[4] P. D'Urso, L.D. Giovanni, and R. Massari, *Time series clustering by a robust autoregressive metric with application to air pollution*, Chemometr. Intell. Lab. Syst. 141 (2015), pp. 107–124.

[5] P. D'Urso and E.A. Maharaj, *Autocorrelation-based fuzzy clustering of time series*, Fuzzy Sets Syst. 160 (2009), pp. 3565–3589.

[6] M. Dash and H. Liu, *Feature selection for classification*, Intell. Data Anal. 1 (1997), pp. 131–156.

[7] B.F. De Blasio, M. Laane, T. Walmann, and I. Giaever, *Combining optical and electrical impedance techniques for quantitative measurement of confluence in MDCK-I cell cultures*, BioTechniques 36 (2004), pp. 650–662.

[8] S. Fekri-Ershad, *Cell phenotype classification using multi threshold uniform local ternary patterns in fluorescence microscope images*, Multimed. Tools. Appl. 80 (2021), pp. 12103–12116.

[9] P. Galeano and D. Peña, *Multivariate analysis in vector time series*, Resenhas 4 (2000), pp. 383–404.

[10] M.L. Gelsinger, L.L. Tupper, and D.S. Matteson, *Cell line classification using electric cell-substrate impedance sensing (ECIS)*, Int. J. Biostat. 16 (2019), p. 20180083.

[11] I. Giaever and C.R. Keese, *Micromotion of mammalian cells measured electrically*, Proc. Natl. Acad. Sci. USA 88 (1991), pp. 7896–7900.

[12] I. Heijink, S. Brandenburg, D. Postma, and A. van Oosterhout, *Cigarette smoke impairs airway epithelial barrier function and cell–cell contact recovery*, Eur. Respiratory J. 39 (2012), pp. 419–428.

[13] J. Hong, K. Kandasamy, M. Marimuthu, C.S. Choi, and S. Kim, *Electrical cell-substrate impedance sensing as a non-invasive tool for cancer cell study*, Analyst 136 (2011), pp. 237–245.

[14] P. Hughes, D. Marshall, Y. Reid, H. Parkes, and C. Gelber, *The costs of using unauthenticated, over-passaged cell lines: how much more data do we need*? Biotechniques 43 (2007), pp. 575–582.

[15] E. Keogh, S. Lonardi, C.A. Ratanamahatana, L. Wei, S.H. Lee, and J. Handley, *Compression-based data mining of sequential data*, Data. Min. Knowl. Discov. 14 (2007), pp. 99–129.

[16] M. Kowolenko, C. Keese, D. Lawrence, and I. Giaever, *Measurement of macrophage adherence and spreading with weak electric fields*, J. Immunol. Methods. 127 (1990), pp. 71–77.
[17] D.C. Lovelady, J. Friedman, S. Patel, D.A. Rabson, and C.M. Lo, *Detecting effects of low levels of cytochalasin B in 3T3 fibroblast cultures by analysis of electrical noise obtained from cellular micromotion*, Biosens. Bioelectron. 24 (2009), pp. 2250–2254.
[18] S. Lukic and J. Wegener, *Impedimetric monitoring of cell-based assays*, in *eLS*, John Wiley & Sons, Ltd, Chichester, 2015. doi: 10.1002/9780470015902.a0025710
[19] E.A. Maharaj, *Clusters of time series*, J. Classif. 17 (2000), pp. 297–314.
[20] E.A. Maharaj, P. D'Urso, and J. Caiado, *Time Series Clustering and Classification*, CRC Press, Boca Raton, FL, 2019.
[21] L. Nanni and A. Numini, *A reliable method for cell phenotype image classification*, Artif. Intell. Med. 43 (2008), pp. 87–97.
[22] G. Park, C.K. Choi, A.E. English, and T.E. Sparer, *Electrical impedance measurements predict cellular transformation*, Cell. Biol. Int. 33 (2009), pp. 429–433.
[23] D. Piccolo, *A distance measure for classifying ARIMA models*, J. Time Ser. Anal. 11 (1990), pp. 153–164.
[24] R Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2021). https://www.R-project.org/.
[25] M.J. Rutten, B. Laraway, C.R. Gregory, H. Xie, C. Renken, C. Keese, and K.W. Gregory, *Rapid assay of stem cell functionality and potency using electric cell-substrate impedance sensing*, Stem Cell Res. Therapy 6 (2015), p. 192.
[26] D. Schneider, M. Tarantola, and A. Janshoff, *Dynamics of TGF-β induced epithelial-to-mesenchymal transition monitored by electric cell-substrate impedance sensing*, Biochim. Biophys. Acta. 1813 (2011), pp. 2099–2107.
[27] J.A. Stolwijk, C. Hartmann, P. Balani, S. Albermann, C.R. Keese, I. Giaever, and J. Wegener, *Impedance analysis of adherent cells after in situ electroporation: non-invasive monitoring during intracellular manipulations*, Biosens. Bioelectron. 26 (2011), pp. 4720–4727.
[28] H. Tong and P. Dabas, *Clusters of time series models: an example*, J. Appl. Stat. 17 (1990), pp. 187–198.
[29] L.L. Tupper, C.R. Keese, and D.S. Matteson, *Time series features for classification of contaminated cell cultures*, MileTS '21: 7th KDD Workshop on Mining and Learning from Time Series, Singapore, 2021.
[30] X. Wang, K. Smith, and R. Hyndman, *Characteristic-based clustering for time series data*, Data. Min. Knowl. Discov. 13 (2006), pp. 335–364.
[31] L. Wang, L. Wang, H. Yin, W. Xing, Z. Yue, M. Guoe, and J. Cheng, *Real-time, label-free monitoring of the cell cycle with a cellular impedance sensing chip*, Biosens. Bioelectron. 25 (2010), pp. 990–995.
[32] H. Wickham, M. Averick, J. Bryan, W. Chang, L.D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T.L. Pedersen, E. Miller, S.M. Bache, K. Müller, J. Ooms, D. Robinson, D.P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, *Welcome to the tidyverse*, J. Open Source Softw. 4 (2019), p. 1686.
[33] W. Zhang, M. Griffin, and D.S. Matteson, *Modeling nonlinear growth followed by long-memory equilibrium with unknown change point*, Submission (2021).