**Growth curve and Substrate consumption graph analysis**

Interfaz de usuario gráfica

Descripción generada automáticamente

**Figure** **.** Workflow of the process and results of the sparse-growth-curve algorithm. **(A)** Plot of the input of the method using cell density with OD600 units versus time *t* in hours. **(B)** Each region on the graph are obtained through a decision tree regression. As we can see, the grey circles and squares are the instantaneous doubling rates calculated from the input data. The red squares represent noise-removed rates versus time for the regression. Finally, it is obtaining a two-step dotted line graph with each step representing a different growth phase. **(C)** The final result of this pipeline is a graph with two areas, the red part represents the exponential phase and the dotted black the linear one. The doubling rates are presented next to each corresponding line. **(D)** The second result is the two-piece linear regression using the RANSAC method with its calculated doubling rate presented next to each phase.

The results of the growth and substrate consumption experiments gave us data in the form of cell density (OD600), substrate concentration in g/L, and time (t) in hours. To automate its analysis, a software package designed for parsing cellular growth curves with low temporal resolution was applied. The sparse-growth-curve is available on GitHub under MIT license (Trash and Cheng, 2022).

The code was modified from the repository to effectively analyze the consumption curve of glucose, acetate, and propionate. Using Anaconda Software Distribution (2020), it was possible to create a Jupyter notebook (Kluyver, T. et al., 2016) containing the code. To run it, the environment was built using a Python 3.8 version (Van Rossum and Drake FL, 1995) and employing several packages such as NumPy v1.19.5 for numerical analysis (Harris et al., 2020), ScipPy v1.4.1 (Virtanen et al., 2020), Scikit-learn v0.22.2 for statistical analysis (Pedregosa et al., 2011), Pandas v1.1.15 for handling table format data (McKinney, 2010) and Matplotlib v3.2.2 for data visualization (Hunter, 2007).

From the data, an array containing the time and OD600 of each measure was created and then plotted (Fig A). Additionally, cell density was scaled logarithmically to clearly obtain each one of the growth phases. Using a decision tree regression model, the growth phases can be predicted by grouping the data points with similar instantaneous growth rate (Figure B). Finally, two lines were drawn delimiting each of the growth phases (exponential and linear) and thus calculating the growth rates using a linear regression (Figure C).

To compare the estimation made with the regression tree, an iterative random sample consensus (RANSAC) was implemented too (Figure D). Using this nondeterministic algorithm, two lines were fit again with the instantaneous growth rate of each point and completely ignoring those that could be consider “outliers” from the general population of samples (G. Derpanis, 2010).

For the analysis of the substrate consumption, the algorithm follows the same workflow with the glaring difference of not using a logarithmic scale.

Bibliography

Anaconda Software Distribution. (2020). Anaconda Documentation. Anaconda Inc. Retrieved from https://docs.anaconda.com/

G. Derpanis, K., 2010. Overview of the RANSAC Algorithm. [online] Available at: <http://rmozone.com/snapshots/2015/07/cdg-room-refs/ransac.pdf> [Accessed 1 June 2022].

Harris, C., Millman, K., van der Walt, S., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M., Brett, M., Haldane, A., del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. and Oliphant, T., 2020. Array programming with NumPy. Nature, 585(7825), pp.357-362.

Hunter, J., 2007. Matplotlib: A 2D Graphics Environment. Computing in Science &amp; Engineering, 9(3), pp.90-95.

Kluyver, T., Ragan-Kelley, B., Fernando Perez, Granger, B., Bussonnier, M., Frederic, J., … Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87–90)

McKinney, W., 2010. Data Structures for Statistical Computing in Python. Proceedings of the Python in Science Conference.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

Trash, C. and Cheng, C., 2022. thrash-lab / sparse-growth-curve. [online] GitHub. Available at: <https://github.com/ thrash-lab/sparse-growth-curve> [Accessed 1 June 2022].

Van Rossum, G. and Drake FL, J., 1995. Python tutorial. Centrum voor Wiskunde en Informatica, Amsterdam, Netherlands.. [online] Available at: <https://www.python.org/> [Accessed 1 June 2022].

Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D., Hagen, D., Pasechnik, D., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G., Ingold, G., Allen, G., Lee, G., Audren, H., Probst, I., Dietrich, J., Silterra, J., Webber, J., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J., de Miranda Cardoso, J., Reimer, J., Harrington, J., Rodríguez, J., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P., Lee, P., McGibbon, R., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T., Robitaille, T., Spura, T., Jones, T., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. and Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), pp.261-272.

Annexes. Growth and Substrate Consumption Rates obtain from the Sparse-Growth-Curve Algorithm

Growth Rate

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Strain** |  | **Phase I (0 - ~18 h)** | **Phase II (~18 - ~28 h)** | **Phase III (~28 - ~50 h)** |
| POd1U-L- |  | 0.38 | 0 | 0 |
| Y2900U+L+(JMN Results)(1) |  | 0.38 | 0.04 | 0.04 |
| Y2900U+L+(JMN Results)(2) |  | 0.32 | 0.04 | 0.04 |
| Y2900U+L+(JMN Results)(3) |  | 0.26 | 0.26 | 0 |
| Y8967mcc1mcc2(1) |  | 0.42 | 0.03 | 0.03 |
| Y8967mcc1mcc2(2) |  | 0.35 | 0.06 | 0.06 |
| Y8967mcc1mcc2(3) |  | 0.21 | 0.21 | 0 |
| Y8968mcc1mcc2(1) |  | 0.39 | 0.03 | 0.03 |
| Y8968mcc1mcc2(2) |  | 0.33 | 0.08 | 0.08 |
| Y8968mcc1mcc2(3) |  | 0.22 | 0.22 | 0.19 |
| Y8969mcc1mcc2MCCmito(1) |  | 0.07 | 0.07 | 0.4 |
| Y8969mcc1mcc2MCCmito(2) |  | 0.26 | 0.26 | 0.7 |
| Y8969mcc1mcc2MCCmito(3) |  | 0.04 | 0.04 | 0.04 |
| Y8970mcc1mcc2MCCmito(1) |  | 0.06 | 0.06 | 0.18 |
| Y8970mcc1mcc2MCCmito(2) |  | 0.16 | 0.19 | 0.19 |
| Y8970mcc1mcc2MCCmito(3) |  | 0.03 | 0.04 | 0.04 |
| Y8971(P) |  | 0.03 | 0.04 | 0.04 |
| Y8971(PA) |  | 0.22 | 0.22 | 0.18 |
| Y8971(PG) |  | 0.06 | 0.06 | 0.41 |
| Y8972(P) |  | 0 | 0.05 | 0.05 |
| Y8972(PA) |  | 0.22 | 0.22 | 0.19 |
| Y8972(PG) |  | 0.06 | 0.06 | 0.45 |
| Y8973(P) |  | 0.05 | 0.05 | 0.05 |
| Y8973(PA) |  | 0.29 | 0.29 | 0.04 |
| Y8973(PG) |  | 0.06 | 0.06 | 0.33 |
| Y8974(P) |  | -0.07 | 0.05 | 0.05 |
| Y8974(PA) |  | 0.33 | 0.09 | 0.09 |
| Y8974(PG) |  | 0.08 | 0.33 | 0.22 |
| Y8975(P) |  | 0 | 0.05 | 0.05 |
| Y8975(PA) |  | 0.28 | 0.05 | 0.03 |
| Y8975(PG) |  | 0.07 | 0.07 | 0.33 |
| Y8976(P) |  | 0.03 | 0.06 | 0.06 |
| Y8976(PA) |  | 0.27 | 0.27 | 0.04 |
| Y8976(PG) |  | 0.08 | 0.08 | 0.33 |
| Y9200(P) |  | 0.32 | 0.32 | 0.02 |
| Y9200(PA) |  | 0.71 | 0.71 | 0.07 |
| Y9200(PG) |  | 0.36 | 0.36 | 0.03 |
| Y9850mcc1mcc2RPCT(1) |  | 0.06 | 0.06 | 0.27 |
| Y9850mcc1mcc2RPCT(2) |  | 0.29 | 0.06 | 0.06 |
| Y9850mcc1mcc2RPCT(3) |  | 0 | 0.03 | 0.05 |
| Y9851mcc1mcc2RPCT(1) |  | 0.06 | 0.06 | 0.33 |
| Y9851mcc1mcc2RPCT(2) |  | 0.3 | 0.3 | 0.02 |
| Y9851mcc1mcc2RPCT(3) |  | 0 | 0.05 | 0.05 |

Glucose consumption Rate

|  |  |  |  |
| --- | --- | --- | --- |
| **Strain** | **Phase I (0 - ~18 h)** | **Phase II (~18 - ~28 h)** | **Phase III (~28 - ~50 h)** |
| POd1U-L- | 0 | -0.07 | -0.07 |
| Y2900U+L+(JMN Results)(1) | -0.02 | -0.39 | 0 |
| Y8967mcc1mcc2(1) | -0.01 | -0.51 | 0 |
| Y8968mcc1mcc2(1) | 0 | -0.4 | 0 |
| Y8969mcc1mcc2MCCmito(1) | 0 | 0 | -0.13 |
| Y8970mcc1mcc2MCCmito(1) | 0 | 0 | -0.03 |
| Y8971(PG) | 0 | 0 | -0.05 |
| Y8972(PG) | 0 | 0 | -0.07 |
| Y8973(PG) | 0.06 | -0.07 | -0.24 |
| Y8974(PG) | 0 | 0 | -0.19 |
| Y8975(PG) | 0 | 0 | -0.04 |
| Y8976(PG) | 0 | 0 | -1.18 |
| Y9200(PG) | -0.01 | -0.24 | -0.24 |
| Y9850mcc1mcc2RPCT(1) | -0.01 | -0.01 | -0.07 |
| Y9851mcc1mcc2RPCT(1) | 0 | 0 | -0.21 |

Propanoate Consumption Rate

|  |  |  |  |
| --- | --- | --- | --- |
| **Strain** | **Phase I (0 - ~18 h)** | **Phase II (~18 - ~28 h)** | **Phase III (~28 - ~50 h)** |
| POd1U-L- | -0.02 | -0.02 | -0.02 |
| Y2900U+L+(JMN Results)(1) | 0 | -0.04 | -0.16 |
| Y2900U+L+(JMN Results)(2) | -0.01 | -0.1 | -0.11 |
| Y2900U+L+(JMN Results)(3) | 0 | -0.14 | -0.14 |
| Y8967mcc1mcc2(1) | 0 | -0.06 | -0.18 |
| Y8967mcc1mcc2(2) | -0.01 | -0.11 | -0.11 |
| Y8967mcc1mcc2(3) | 0 | -0.2 | -0.3 |
| Y8968mcc1mcc2(1) | 0 | -0.06 | -0.11 |
| Y8968mcc1mcc2(2) | 0 | -0.02 | -0.28 |
| Y8968mcc1mcc2(3) | 0 | -0.22 | -0.24 |
| Y8969mcc1mcc2MCCmito(1) | -0.02 | -0.02 | -0.2 |
| Y8969mcc1mcc2MCCmito(2) | 0 | -0.05 | -0.05 |
| Y8969mcc1mcc2MCCmito(3) | 0 | 0 | 0 |
| Y8970mcc1mcc2MCCmito(1) | 0 | 0 | 0 |
| Y8970mcc1mcc2MCCmito(2) | -0.01 | -0.01 | -0.08 |
| Y8970mcc1mcc2MCCmito(3) | 0 | 0 | 0 |
| Y8971(P) | 0 | 0 | 0 |
| Y8971(PA) | 0 | 0 | -0.03 |
| Y8971(PG) | 0 | 0 | -0.03 |
| Y8972(P) | 0 | 0 | 0 |
| Y8972(PA) | 0 | -0.03 | -0.03 |
| Y8972(PG) | 0 | 0 | -0.3 |
| Y8973(P) | 0 | 0 | 0 |
| Y8973(PA) | 0 | 0 | -0.03 |
| Y8973(PG) | 0 | 0 | -0.01 |
| Y8974(P) | 0.01 | 0 | 0 |
| Y8974(PA) | -0.01 | -0.01 | -0.04 |
| Y8974(PG) | 0 | 0 | -0.02 |
| Y8975(P) | 0 | 0 | 0 |
| Y8975(PA) | 0 | -0.03 | -0.03 |
| Y8975(PG) | 0 | -0.03 | -0.03 |
| Y8976(P) | 0 | 0 | 0 |
| Y8976(PA) | 0 | 0 | -0.06 |
| Y8976(PG) | 0 | 0 | -0.03 |
| Y9200(P) | -0.01 | -0.03 | -0.41 |
| Y9200(PA) | -0.01 | -0.01 | -0.1 |
| Y9200(PG) | 0 | -0.03 | -1.6 |
| Y9850mcc1mcc2RPCT(1) | 0 | 0 | -0.02 |
| Y9850mcc1mcc2RPCT(2) | -0.01 | -0.04 | -0.04 |
| Y9850mcc1mcc2RPCT(3) | 0 | 0 | 0 |
| Y9851mcc1mcc2RPCT(1) | 0 | 0 | -0.04 |
| Y9851mcc1mcc2RPCT(2) | 0 | -0.04 | -0.04 |
| Y9851mcc1mcc2RPCT(3) | 0 | 0 | 0 |

Acetate Consumption Rate

|  |  |  |  |
| --- | --- | --- | --- |
| **Strain** | **Phase I (0 - ~18 h)** | **Phase II (~18 - ~28 h)** | **Phase III (~28 - ~50 h)** |
| Y2900U+L+(JMN Results)(2) | -0.01 | -0.63 | 0 |
| Y8967mcc1mcc2(2) | 0 | -0.26 | -0.26 |
| Y8968mcc1mcc2(2) | 0 | -0.21 | 0 |
| Y8969mcc1mcc2MCCmito(2) | 0 | -0.1 | -0.09 |
| Y8970mcc1mcc2MCCmito(2) | 0 | -0.11 | -0.16 |
| Y8971(PA) | -0.01 | -0.01 | -0.12 |
| Y8972(PA) | -0.02 | -0.16 | -0.16 |
| Y8973(PA) | -0.01 | -0.09 | -0.09 |
| Y8974(PA) | 0 | -0.04 | -0.24 |
| Y8975(PA) | -0.03 | -0.03 | -0.2 |
| Y8976(PA) | 0 | -0.05 | -0.4 |
| Y9200(PA) | -0.01 | -0.43 | -0.43 |
| Y9850mcc1mcc2RPCT(2) | 0 | -0.244 | -0.244 |
| Y9851mcc1mcc2RPCT(2) | 0 | -0.47 | -0.47 |