Identification and Prediction of Flux Tower Latent Heat Data and Their Source Variables (Time Series Imputation)

This manuscript ([permalink]) was automatically generated from [uiceds/project-team-wres@4b3e5ca] on September 26, 2024.

Authors

- **Jiaze Cao** ✉
  · JiazeCLeo
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; WRES

- **Yuyao Huang**
  · Yuyao-Huang
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; EWES

- **Yue Wan**
  · clarawan
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; EWES · Funded by Grant

- **Hsing-Yu Huang** ✉
  · Hsing-Yu
  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign; WRES

✉ — Correspondence possible via GitHub Issues or email to Jiaze Cao <jiazec2@illinois.edu>, Hsing-Yu Huang <hsingyu3@illinois.edu>.

Dataset Description

We propose to use Goose Creek Eddy Covariance Flux Tower Sensor Data[**kumar2024?**]. The data is collected from the Eddy Covariance Flux Tower in Goose Creek, Piatt County. The dataset consists of time series data spanning from Spring 2016 to Spring 2023 with 15 minutes time interval. Dataset involves 167 variables shown in Table 1 including latent heat, sensible heat, wind speed, temperature, and changes in the ecosystem with respect to water, carbon, and temperature. Figure 1 illustrates part of variables in 2022. The data collected by flux tower provides a foundation for further investigation into hydrological, meteorological, and environmental phenomena. The format of dataset is CSV file (generated from raw PICKLE file). The dataset can be found through link: https://www.hydroshare.org/resource/c276c71e8d1246e29d8502f5b2054668/

Fig 1: Variables Latent Heat, Sensible Heat, Air Pressure and Vapor Pressure in 2022

Variable Name	Units	Description
TIMESTAMP	TS	
RECORD	RN	
Hs	W/m^2	sensible heat flux
tau	$kg/(m \cdot s^2)$	shear stress
u_star	m/s	friction velocity
Ts_stdev	$^\circ C$	instantaneous stdev of temperature

Variable Name	Units	Description
Ts_Ux_cov	$^{\circ}C \cdot m/s$	inst. cov(temp, Ux)
Ts_Uy_cov	$^{\circ}C \cdot m/s$	inst. cov(temp, Uy)
Ts_Uz_cov	$^{\circ}C \cdot m/s$	inst. cov(temp, Uv)
Ux_stdev	m/s	instantaneous stdev of Ux
Ux_Uy_cov	$(m/s)^2$	instantaneous cov of (Ux, Uy)
Ux_Uz_cov	$(m/s)^2$	instantaneous cov of (Ux, Uz)
Uy_stdev	m/s	instantaneous stdev of Uy
Uy_Uz_cov	$(m/s)^2$	instantaneous cov of (Uy, Uz)
Uz_stdev	m/s	instantaneous stdev of Uz
wnd_spd	m/s	wind speed (horizontal) - different from next?
rslt_wnd_spd	m/s	wind speed (horizontal)
wnd_dir_sonic	$degrees$	wind direction from CSAT3, deg from N?
std_wnd_dir	$degrees$	inst. stdev of wind direction
wnd_dir_compass	$degrees$	wind direction from compass (from N?)
Ux_Avg	m/s	average horiz windspeed x
Uy_Avg	m/s	average horiz windspeed y
Uz_Avg	m/s	average vertical windspeed z
Ts_Avg	$^{\circ}C$	air temperature at 25 m
sonic_azimuth	$degrees$	180 is direction is pointing - can change this value
sonic_samples_Tot	samples	10 Hz sampling rate (cycles per 15 mins = 9000)
Fc_li_wpl	$mg/(m^2 \cdot s)$	carbon flux upward (+ = upward) with Webb et al Term
LE_li_wpl	W/m^2	latent heat flux with Webb et al term
Hc_li	W/m^2	sensible heat flux
CO2_li_mean	mg/m^3	CO2 conc –> need to convert to ppm units
H2O_li_mean	g/m^3	water vapor conc at 25 m
amb_press_li_mean	kPa	air pressure at 25 m
Tc_li_mean	$^{\circ}C$	CSAT air temperature at 25 m
rho_a_li_mean	kg/m^3	density of air with water vapor
Fc_li_irga	$mg/(m^2 \cdot s)$	carbon flux without Webb et al. Term
LE_li_irga	W/m^2	latent heat flux without Webb et al. Term
irga_li_samples_Tot	$samples$	should be around 60 - quality indicator of LiCor
Precip_Tot	mm	rainfall

Variable Name	Units	Description
T_tmpr_rh_mean	$^{\circ}C$	air temperature at 25 m
e_tmpr_rh_mean	kPa	vapor pressure at 25 m
e_sat_tmpr_rh_mean	kPa	saturated vapor pressure at 25 m
H2O_tmpr_rh_mean	g/m^3	water vapor conc at 25 m
RH_tmpr_rh_mean		Relative Humidity at 25 m (e/e_sat)
rho_a_tmpr_rh_mean	kg/m^3	air density
slowsequence_1_Tot	$samples$	cycles per 15 mins - scanning every 10 secs
CS655_Wcr_Avg	m^3/m^3	soil water content
CS655_Ec_Avg	dS/m	soil conductivity
CS655_Tmpr_Avg	$^{\circ}C$	soil temperature
mean_wind_speed	m/s	wind speed at 10 m height...not average?
mean_wind_direction	$degrees$	wind direction
std_wind_dir	$degrees$	mean wind vector stdev of direction
NDVI_Avg		Normalized Difference Vegetation Index
NDVIUpRed_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIUpNIR_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIIndUp		NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIDownRed_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIDownNIR_Avg	$W/m^2 \cdot nm$	NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
NDVIIndDown		NDVI is calculated from upward and canopy facing sensors that measure IR and NIR radiation
PRI_Avg		Photochemical Reflectance Index
PRIUp531_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIUp570_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIIndUp		PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation

Variable Name	Units	Description
PRIDown531_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIDown570_Avg	$W/m^2 \cdot nm$	PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
PRIIndDown		PRI calculated from updward and canopy facing sensors that measure 2 wavelengths of radiation
D5TE_VWC_5cm_Avg	m^3/m^3	volumetric water content
D5TE_P_5cm_Avg		bulk dielectric permittivity
D5TE_EC_5cm_Avg	dS/m	soil electrical conductivity
D5TE_T_5cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_15cm_Avg	m^3/m^3	volumetric water content
D5TE_P_15cm_Avg		bulk dielectric permittivity
D5TE_EC_15cm_Avg	dS/m	soil conductivity
D5TE_T_15cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_30cm_Avg	m^3/m^3	volumetric water content
D5TE_P_30cm_Avg		bulk dielectric permittivity
D5TE_EC_30cm_Avg	dS/m	soil conductivity
D5TE_T_30cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_50cm_Avg	m^3/m^3	volumetric water content
D5TE_P_50cm_Avg		bulk dielectric permittivity
D5TE_EC_50cm_Avg	dS/m	soil conductivity
D5TE_T_50cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_100cm_Avg	m^3/m^3	volumetric water content
D5TE_P_100cm_Avg		bulk dielectric permittivity
D5TE_EC_100cm_Avg	dS/m	soil conductivity

Variable Name	Units	Description
D5TE_T_100cm_Avg	$^{\circ}C$	soil temperature
D5TE_VWC_200cm_Avg	m^3/m^3	volumetric water content
D5TE_P_200cm_Avg		bulk dielectric permittivity
D5TE_EC_200cm_Avg	dS/m	soil conductivity
D5TE_T_200cm_Avg	$^{\circ}C$	soil temperature
slowsequence_2_Tot	samples	cycles - 1 minute loops (number of times scanned)
SB121TempC_Avg	$^{\circ}C$	SB = sensor body, temp of body of sensor
Targ121TempC_Avg	$^{\circ}C$	surface temperature
Targ121mV_Avg	$^{\circ}C$	
SB1H1TempC_Avg	$^{\circ}C$	SB = sensor body, temp of body of sensor
Targ1H1TempC_Avg	$^{\circ}C$	surface temperature
Targ1H1mV_Avg	$^{\circ}C$	
short_up_Avg	W/m^2	Incoming shortwave radiation detected by the upward facing instrument
short_dn_Avg	W/m^2	Outgoing shortwave radiation detected by the downward facing instrument
long_up_Avg	W/m^2	incoming longwave radiation detected by upward facing instrument
long_dn_Avg	W/m^2	outgoing longwave radiation detected by downward facing instrument
cnr4_T_C_Avg	$^{\circ}C$	temperature of sensor
cnr4_T_K_Avg	K	temperature of sensor in Kelvin
long_up_corr_Avg	W/m^2	Incoming longwave radiation detected by the upward facing instrument, corrected
long_dn_corr_Avg	W/m^2	Outgoing longwave radiation detected by the downward facing instrument, corrected
Rs_net_Avg	W/m^2	Shortwave net radiation (Rshort_up - Rshort_down)
Rl_net_Avg	W/m^2	Longwave net radiation (Rlong_up - Rlong_down)
albedo_Avg	W/m^2	Albedo
Rn_Avg	W/m^2	Net radiation (Rs_net + Rl_net)
SQ_110_Avg	μmol photons $m^{-2}s^{-1}$	PAR (photosynthetically active radiation)
shf_Avg(1)	W/m^2	Ground heat flux

Variable Name	Units	Description
shf_Avg(2)	W/m^2	Ground heat flux
slowsequence_3_Tot	$samples$	number of times scanned in 15 mins (once per min)

Table: A table with a variables.

{#tbl:bowling-scores}

Proposal

Background

Evapotranspiration (ET) is the process of water transferring from land to the atmosphere, accompanying the phase change of water from liquid to gas. This process plays a critical role in the ecohydrological system and profoundly affects the hydrological cycle. The processes of evapotranspiration and energy exchange are interdependent. Both latent heat (LE) and evapotranspiration (ET), from the perspective of energy and water flux, are key terms for anticipating weather conditions, simulating climate, and diagnosing climate change. However, the measurement of evapotranspiration is challenging because the process itself is invisible and complex.

Figure 2 shows the latent heat data gap in 2020 due to covid-19 and overhaul of equipment. Our project goal is to fill in these missing data. The ground truth data is collected from satelite sensors (https://etdata.org/). Despite the existence of numerous classical evapotranspiration simulation models, such as Bowen Ratio, Priestley-Taylor and Penman-Monteith models, the predictive accuracy of these models is inferior to that of deep learning models. Therefore, we plan to use RNN and LSTM deep learning models to predict latent heat and fill the gap.

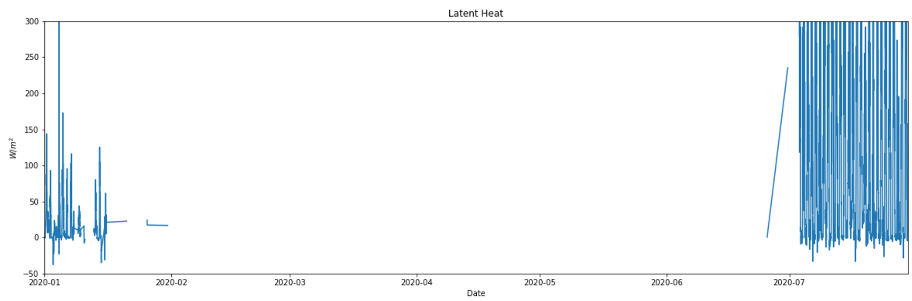

Fig 2: Data Gap in 2020

Step 1: Regression analysis

We have 167 variables in the dataset. Although we can filter some ET related variables based on empirical models, these variables may not accurate and AI models tend to obtain adequate information. Therefore, we propose to conduct regression analysis to find out variables highly correlated to latent heat. These variables will be input variables in deep learning model.

Step 2: Deep Learning Time Series Forecast (Time Series Imputation)

Once we confirm the input variables, we plan to use RNN or LSTM forecast models to predict latent heat in 2020. All the input are divided into training datasets and the validation datasets. After the RNN model is trained, the validation datasets are used to verify the model. At last, the missing data are generated by the model.

Source

This manuscript is a template (aka “rootstock”) for Manubot, a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input (`.md` files in the `/content` directory) to the output you see below.

Basic formatting

Bold text

Semi-bold text

Centered text

Right-aligned text

Italic text

Combined *italics and **bold***

~~Strikethrough~~

1. Ordered list item
2. Ordered list item
    a. Sub-item
    b. Sub-item
        i. Sub-sub-item
3. Ordered list item
    a. Sub-item

- List item
- List item
- List item

subscript: H_2O is a liquid

superscript: 2^{10} is 1024.

unicode superscripts^{0123456789}

unicode subscripts$_{0123456789}$

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to editing and version control.

Line break without starting a new paragraph by putting
two spaces at end of line.

Document organization

Document section headings:

Heading 1

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6

A heading centered on its own printed page

Horizontal rule:

`Heading 1`'s are recommended to be reserved for the title of the manuscript.

`Heading 2`'s are recommended for broad sections such as *Abstract*, *Methods*, *Conclusion*, etc.

`Heading 3`'s and `Heading 4`'s are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [1].

Citation by PubMed Central ID [2].

Citation by PubMed ID [3].

Citation by Wikidata ID [4].

Citation by ISBN [5].

Citation by URL [6].

Citation by alias [7].

Multiple citations can be put inside the same set of brackets [1,5,7]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [2,3,7,8].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 1

Figure 2

Figure 3

Figure 4

Table 1

Equation 1

Equation 2

Quotes and code

> Quoted text

> Quoted block of text
>
> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```python
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures

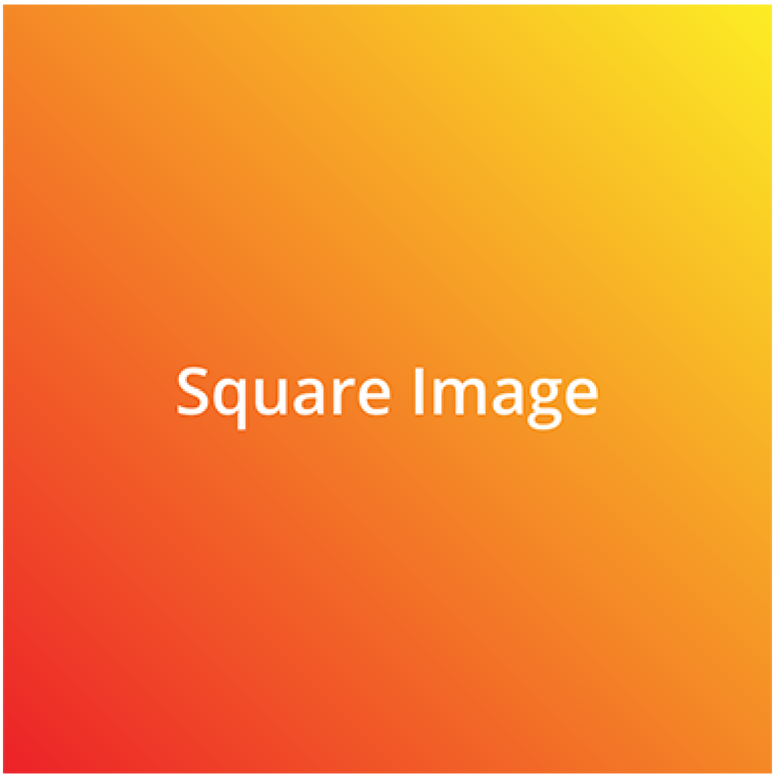

Figure 1: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

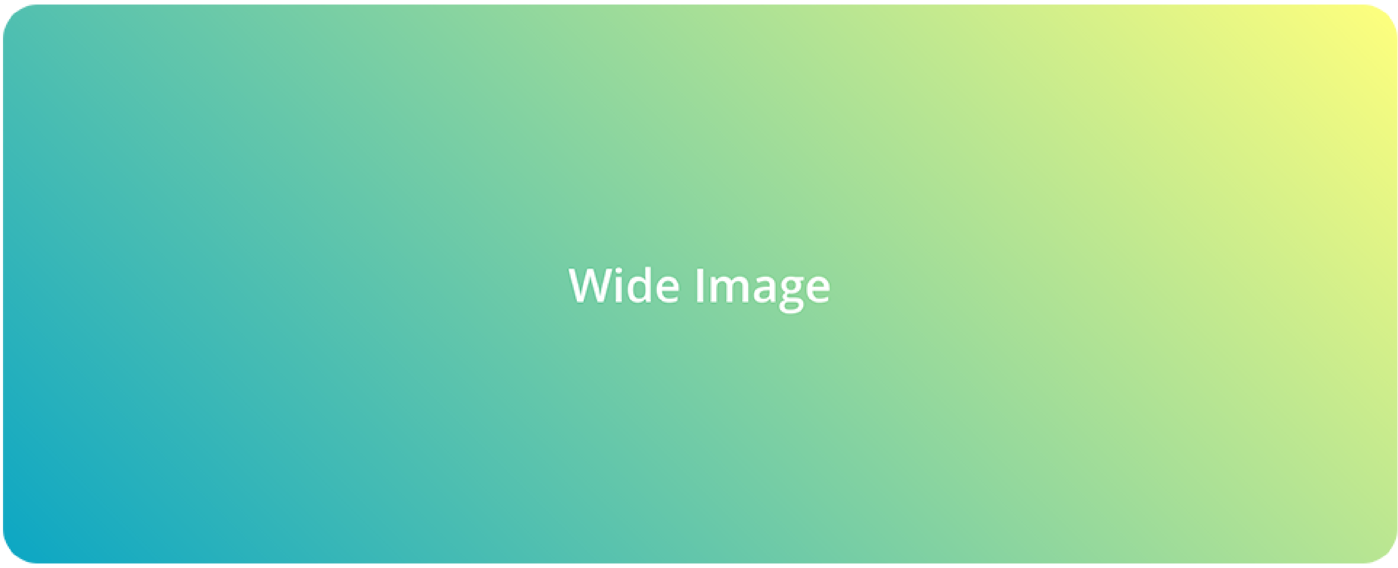

Figure 2: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.

Figure 3: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.

Figure 4: A vector `.svg` image loaded from GitHub. The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image. Note that if you want to export to Word (`.docx`), you need to download the image and reference it locally (e.g. `content/images/vector.svg`) instead of using a URL.

Tables

Table 1: A table with a top caption and specified relative column widths.

Bowling Scores	**Jane**	**John**	**Alice**	**Bob**
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 2: A table too wide to fit within page.

	Digits 1-33	**Digits 34-66**	**Digits 67-99**	**Ref.**
pi	3.14159265358979323 846264338327950	28841971693993751 0582097494459230	78164062862089986 2803482534211706	`piday.org`
e	2.71828182845904523 536028747135266	24977572470936999 5957496696762772	40766303535475945 7138217852516642	`nasa.gov`

Table 3: A table with merged cells using the `attributes` plugin.

	Colors	
Size	**Text Color**	**Background Color**
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{1}$$

An equation too long to fit within page:

$$\begin{aligned} x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \end{aligned} \tag{2}$$

Special

⚠ **WARNING** *The following features are only supported and intended for* `.html` *and* `.pdf` *exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as* `.docx`.

LINK STYLED AS A BUTTON

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

`white` `lightgrey` `grey` `darkgrey` `black` `lightred` `lightyellow` `lightgreen` `lightblue` `lightpurple` `red` `orange` `yellow` `green` `blue` `purple`

Using the [Font Awesome](#) icon set:

✔ ? ★ 🔔 ⊗ ⋯

Light Grey Banner
useful for *general information* - manubot.org

Blue Banner
useful for *important information* - manubot.org

Light Red Banner
useful for *warnings* - manubot.org

References

1. **Sci-Hub provides access to nearly all scholarly literature**
Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene
eLife (2018-03-01) https://doi.org/ckcj
DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410

2. **Reproducibility of computational workflows is automated using continuous analysis**
Brett K Beaulieu-Jones, Casey S Greene
Nature biotechnology (2017-04) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103790/
DOI: 10.1038/nbt.3780 · PMID: 28288103 · PMCID: PMC6103790

3. **Bitcoin for the biological literature.**
Douglas Heaven
Nature (2019-02) https://www.ncbi.nlm.nih.gov/pubmed/30718888
DOI: 10.1038/d41586-019-00447-9 · PMID: 30718888

4. **Plan S: Accelerating the transition to full and immediate Open Access to scientific publications**
cOAlition S
(2018-09-04) https://www.wikidata.org/wiki/Q56458321

5. **Open access**
Peter Suber
MIT Press (2012)
ISBN: 9780262517638

6. **Open collaborative writing with Manubot**
Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter
Manubot (2020-05-25) https://greenelab.github.io/meta-review/

7. **Opportunities and obstacles for deep learning in biology and medicine**
Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, … Casey S Greene
Journal of The Royal Society Interface (2018-04) https://doi.org/gddkhn
DOI: 10.1098/rsif.2017.0387 · PMID: 29618526 · PMCID: PMC5938574

8. **Open collaborative writing with Manubot**
Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter
PLOS Computational Biology (2019-06-24) https://doi.org/c7np
DOI: 10.1371/journal.pcbi.1007128 · PMID: 31233491 · PMCID: PMC6611653