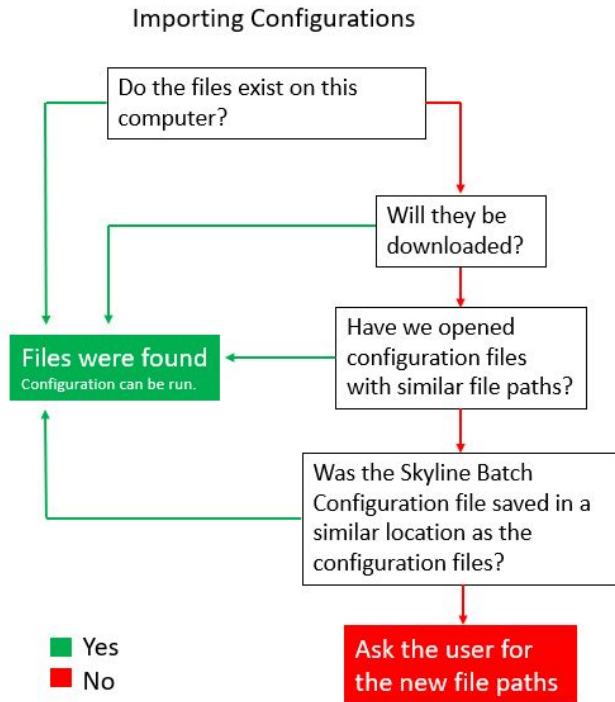# Skyline Batch: An Intuitive User Interface for Batch Processing with Skyline

Authors: Alexandra N. Marsh[1], Vagisha Sharma[2], Surya K. Mani[3], Olga Vitek[4], Michael J. MacCoss[5], Brendan X MacLean*[6]

University of Washington, Genome Sciences 98195[1], University of Washington, Genome Sciences 98195[2], University of Washington, Genome Sciences 98195[3], Northeastern University, Khoury College of Computer Sciences 02115[4], University of Washington, Genome Sciences 98195[5], University of Washington, Genome Sciences 98195[6]

brendanx@u.washington.edu

Supplementary Figure 1. The Skyline Batch algorithm to update file paths when a configuration is opened on a computer, enabling easy porting of configurations between computers
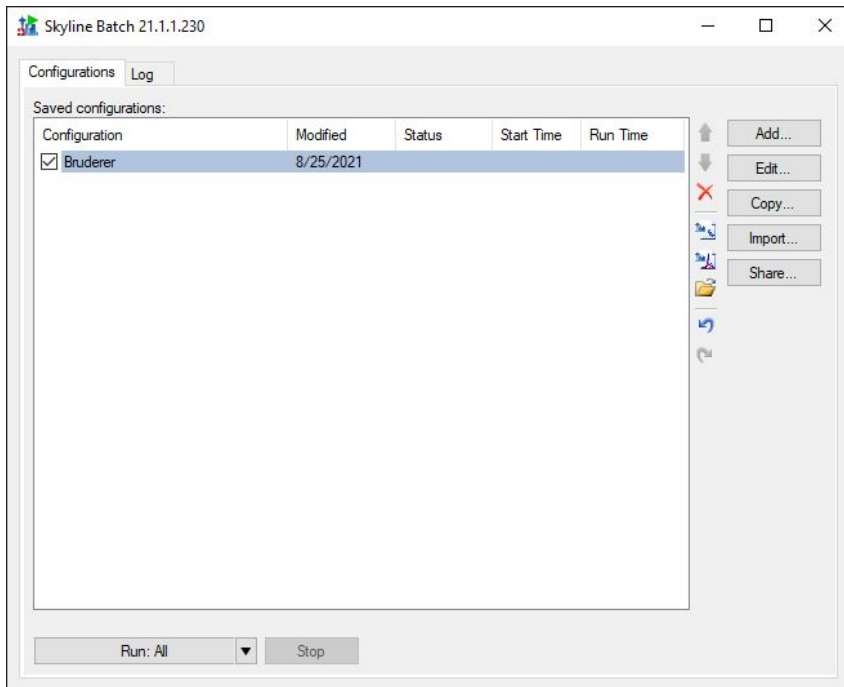
Supplementary Figure 2. A screenshot of Skyline Batch after the Bruderer configuration has been imported from a Skyline Batch configuration file.

Did you successfully run the Bruderer Skyline Batch configuration?   `0` `8` `7`

Yes
31 %

No
69 %

Supplementary Figure 3. A poll taken at the May Institute 2021 course, indicating that one third of respondents successfully ran the Bruderer configuration in Skyline Batch

What is your scripting (R and/or batch) expertise?   `0` `9` `4`

I don't use scripts
41 %

I use scripts occasionally
34 %

I use scripts a lot
10 %

I write scripts and use them
15 %

Supplementary Figure 4. A poll taken at the May Institute 2021 course, indicating that many Skyline users do not have much experience with using and writing scripts

Supplementary Note 1. A link to Webinar 20: Using Skyline Batch for Large-Scale DIA. A webinar fully-focused on Skyline Batch.

https://skyline.ms/webinar20.url

Supplementary Note 2. A link to Webinar 14: Large Scale DIA with Skyline, that has been updated with instructions on running the same analysis in Skyline Batch.

https://skyline.ms/webinar14.url

Supplementary Note 3. A link to Webinar 15: Optimizing Large Scale DIA with Skyline, that has been updated with instructions on running the same analysis in Skyline Batch.

https://skyline.ms/webinar15.url

Supplementary Note 4. The Skyline Batch documentation. For the most up to date version of the documentation, please view it at https://skyline.ms/batch-docs.url.

Skyline Batch

Skyline Batch is a user-friendly tool to run a simple Skyline workflow on many files at a time. When run, it uses an invisible instance of Skyline to do the following:
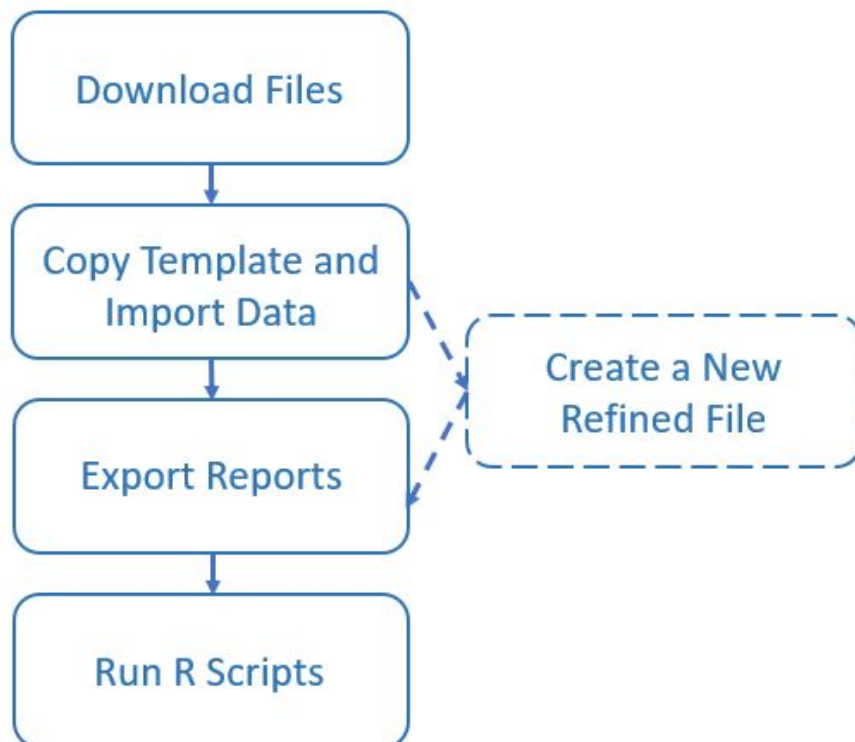
Download Files: If any of the files used during the Skyline Batch run are located on a server, Skyline Batch will check if they exist on disk. If they are missing, they are downloaded during this step.

Copy Template and Import data: This step uses a Skyline template file to create a new Skyline results file. First, the template is copied to a specified "analysis" folder, then data is imported into it. Since the template is copied into a new folder first, the original Skyline template file is not changed and can be used for many different Skyline Batch runs.

Create a New Refined File: This step is only executed if refinement settings are specified (to learn more about adding refinement settings, see the documentation for the **Refine** tab in the **Skyline Batch Configuration** form on page 7). The refinement will be done using the Skyline results file generated in the previous step, then saved as a new file in a specified location. Creating the refined file in this step does not change the Skyline results file or the Skyline template file.

Export reports: Reports are CSV files containing useful information about the data in a Skyline file. They can be created from either the results file or the refined file (if it exists). Configurations do not need to export any reports (although this is necessary if you want to run R scripts). If the configuration has no reports, the Skyline Batch run will finish after the previous step.

Run R Scripts: R scripts are run using the exported reports from the previous step. Each R script receives the data from a specific exported report as input. Running R scripts is not required, you can export reports without running any R scripts on it.

STARTUP

Skyline Batch requires an installation of Skyline (or Skyline-Daily).
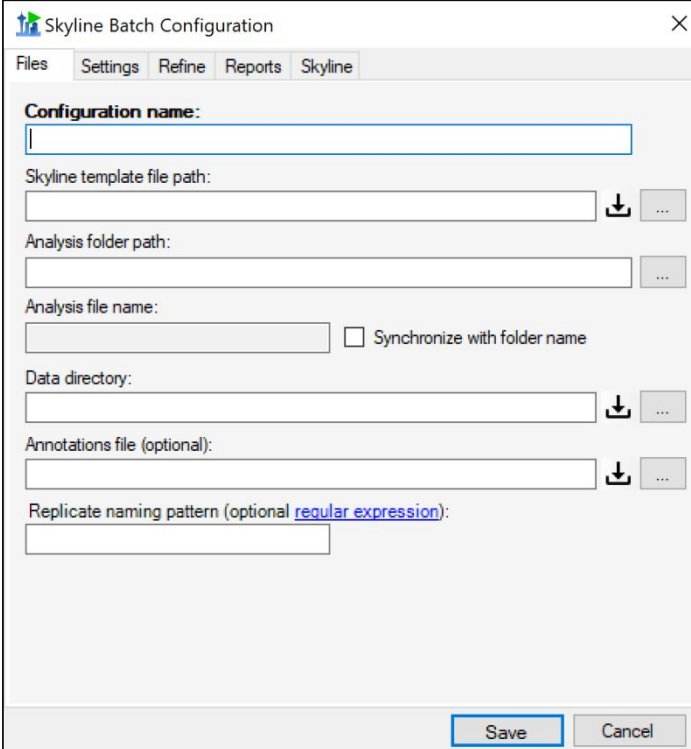

On startup, you will see the main window:
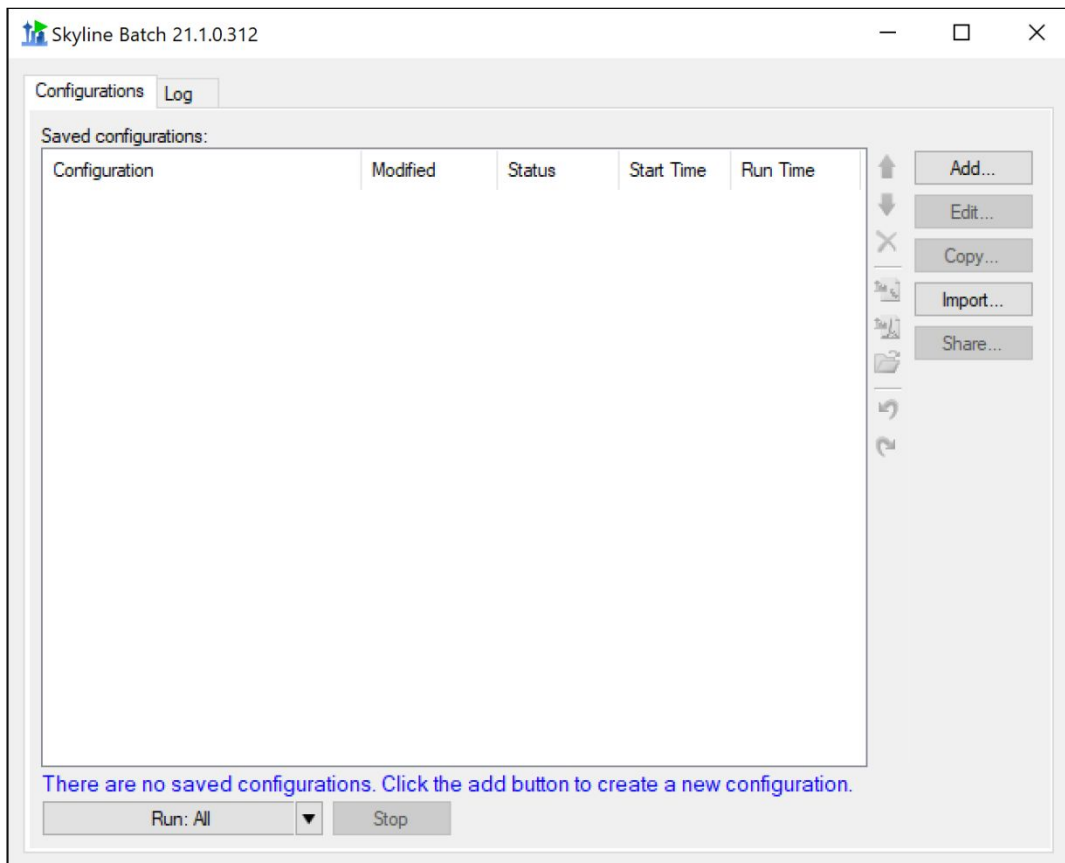
To start using Skyline Batch, create a new configuration by clicking the **Add** button. This brings up the **Skyline Batch Configuration** form.

CONFIGURATION FILES

In the **Skyline Batch Configuration** form enter the appropriate settings in the **Files** tab.

Configuration name: This can be the name of the project or study for which data is being acquired. The configuration name must be unique.

Skyline template file path: This is the path to the template Skyline document that will be copied to the analysis folder and used to import data into.

If you have other configurations that create refined template files, an arrow will appear on the right side of the template file input, allowing you to either select from these files or choose your own template with the download and browse buttons.
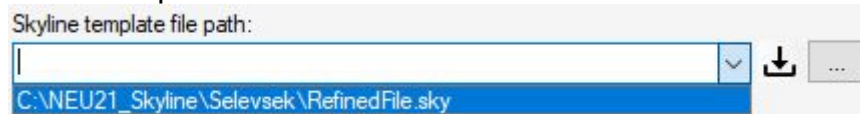
Analysis folder path: The path to the local folder in which the copied template document and exported reports will be stored.

> *Note: For Skyline Batch to find the assay library, the Skyline template file location must be in the parent hierarchy of the analysis folder. (ie: If my Analysis folder is "C:\Users\myaccount\Documents\Analysis", then my template file must be in "users," "myaccount," or "Documents"). This does not apply to refined templates.

Analysis file name: The name of the Skyline file that will be created from the template document in the analysis folder. This will be the same as the name of the Skyline template document or the analysis folder and cannot be directly edited.

Synchronize with folder name: A checkbox dictating whether the new Skyline document will have the same name as the analysis folder (checked) or the same name as the template document (unchecked).
Data directory: The folder containing data to import into the copied template document. This folder should never contain anything aside from the data files you wish to import.
Annotations file: An optional CSV file containing the annotations for the imported data.
Replicate naming pattern: An optional regular expression from which the first group will be used to name replicates when importing.
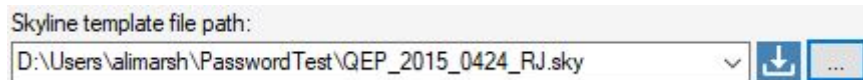
Downloading Files

⬇ This button indicates that the file can be downloaded from a remote server. Its color indicates if a remote server was added, and if the file will be downloaded:

⬇ No server – the file already exists on the computer.

⬇ Server – the file will be downloaded to the specified file path.

For example

Skyline template file path:

D:\Users\alimarsh\PasswordTest\QEP_2015_0424_RJ.sky

The blue server button indicates that QEP_2015_0424_RJ.sky will be downloaded from a server. The Skyline template file path indicates that the downloaded file will be placed in the "D:\Users\alimarsh\PasswordTest\" folder.
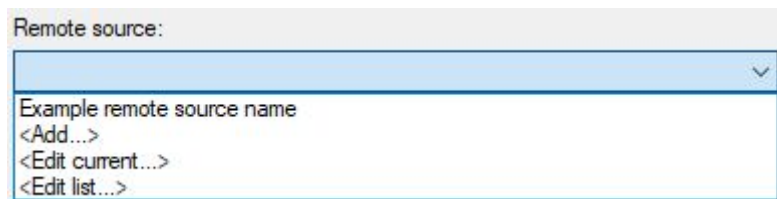
Remote File Sources

Remote file sources simplify the process of downloading files from similar remote locations. They allow users to name a remote source once, this is usually a path to a folder on PanoramaWeb or an FTP server, and then link to files inside that folder using relative paths. Remote file sources are used anywhere a file is downloaded, so they will be covered before the **Downloading Data Files** and **Downloading Panorama Files** sections.

When you open any download file dialog using a download file button (⬇), a **Remote source** combobox is found at the top of the dialog:
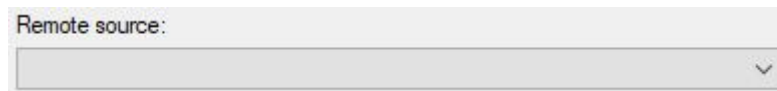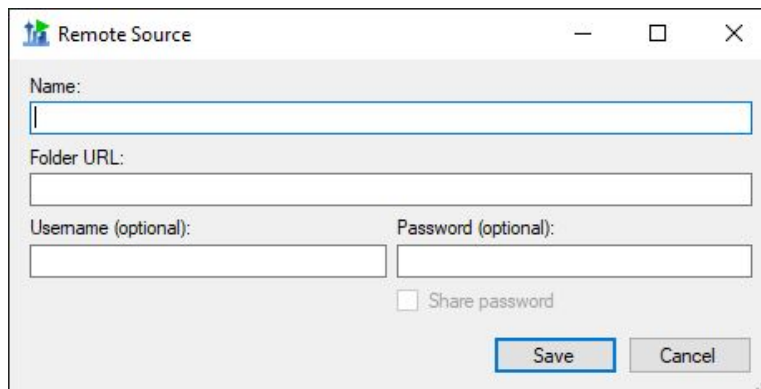
Remote source:

Expanding the combobox gives a list of the existing remote sources, along with options to add, edit current, or edit the list of remote sources:

Remote source:

Example remote source name
<Add...>
<Edit current...>
<Edit list...>

Clicking the <**Edit list…**> option opens the **Edit Remote File Sources** dialog:
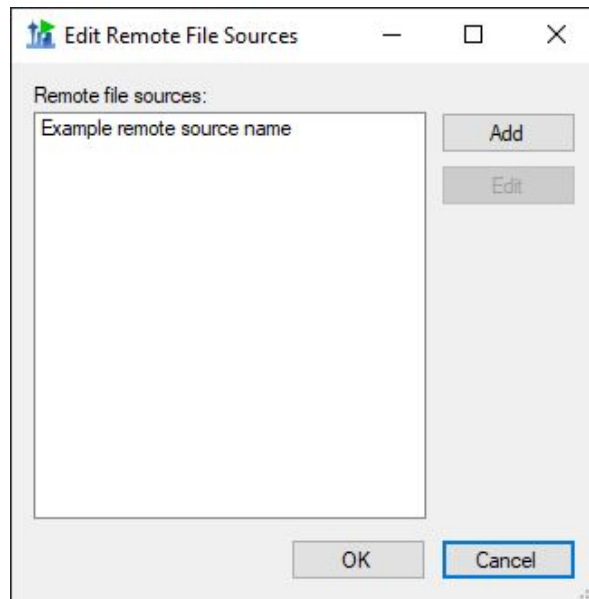
To add a remote file source, click the **Add** button (or choose the **<Add…>** option from the previous combobox). This opens the **Remote Source** dialog:

Name: A unique human-readable name to label the remote file source in the list
URL: The URL to a folder on a Panorama or FTP server. If the URL points to a location on Panorama, it must be the LabKey URL (will have "_webdav" in the URL).
Username: The username for the server if it has one.
**Password**: The password to the server if it has one.
Share Password: If the password should be sharable between computers (stored and saved in plain text instead of being encrypted). This is only recommended for publicly available passwords that you wish to share between computers in a Skyline Batch Configuration file.

When the remote file source is saved, Skyline Batch connects to it and adds it to the list. Remote file sources are selected by choosing them in the **Remote source** combobox.

Downloading Data Files

The download button next to Data directory opens the **Download Remote Data Files** dialog:
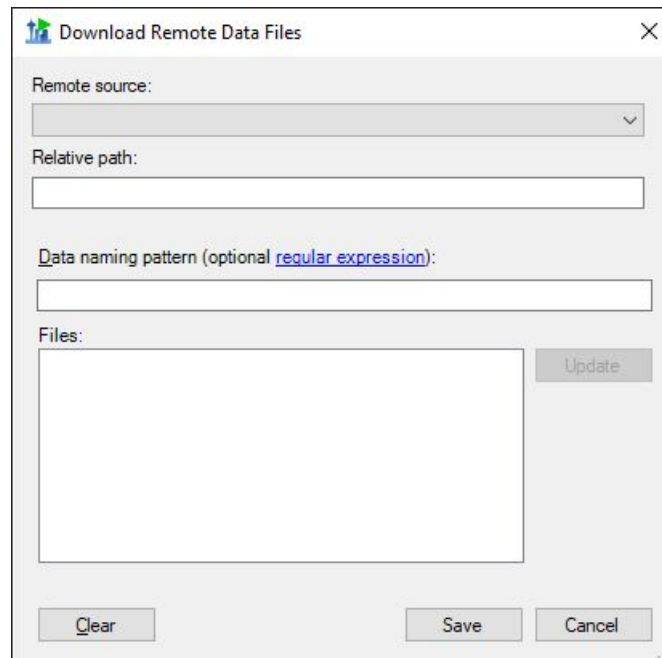
The **Download Remote Data Files** dialog is unique as it allows you to download a set of files from a server, whereas the other download file dialogs will download a single file.
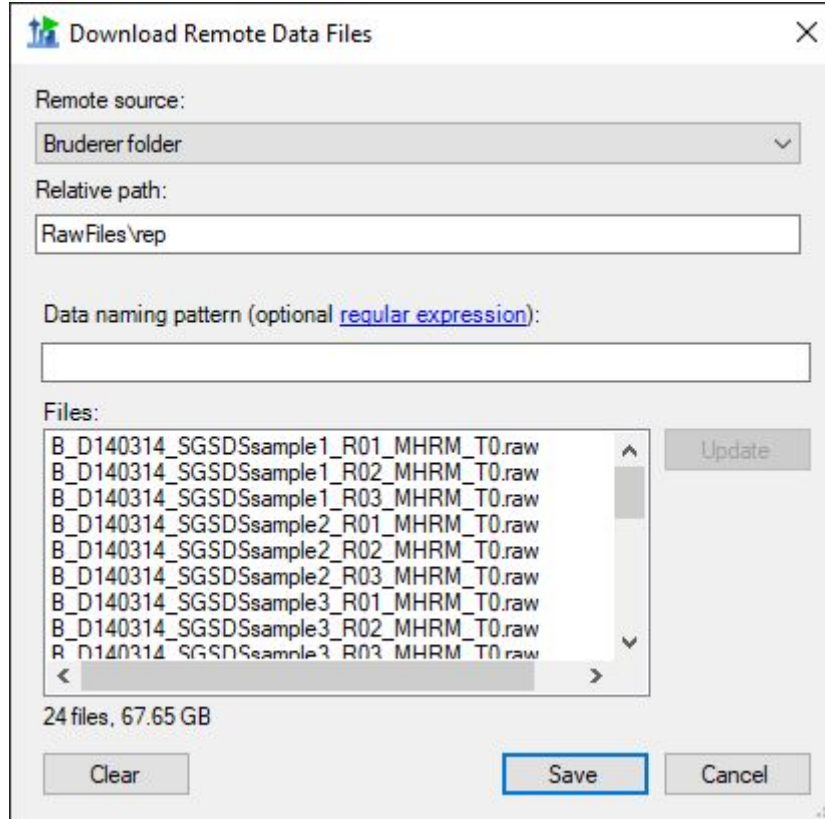
Remote source: The remote source for the data files (see previous section).
Relative path: The relative path to the data folder. This can be empty if the data files are directly inside the folder specified in the remote source.
Data naming pattern (optional): A regular expression that matches the names of the desired data files on the server.

Files: This is a list of the files at the URL that will be downloaded. When the URL, Username, or Password are changed, this list can be updated by clicking the **Update** button. The file list updates automatically if only the regular expression is changed.
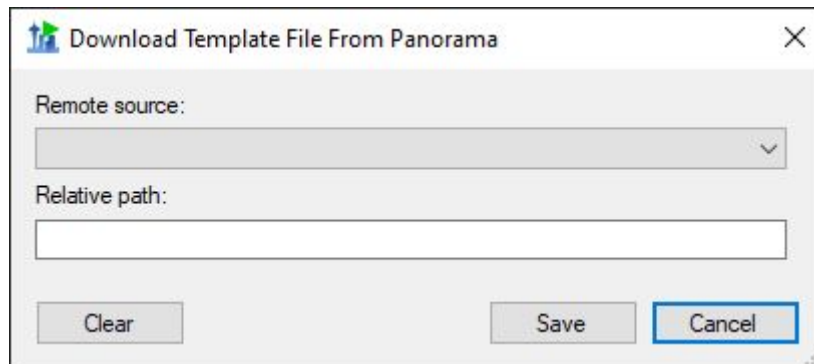
An updated files list looks like this:

The data server can be removed with the **Clear** button.

Downloading Panorama Files

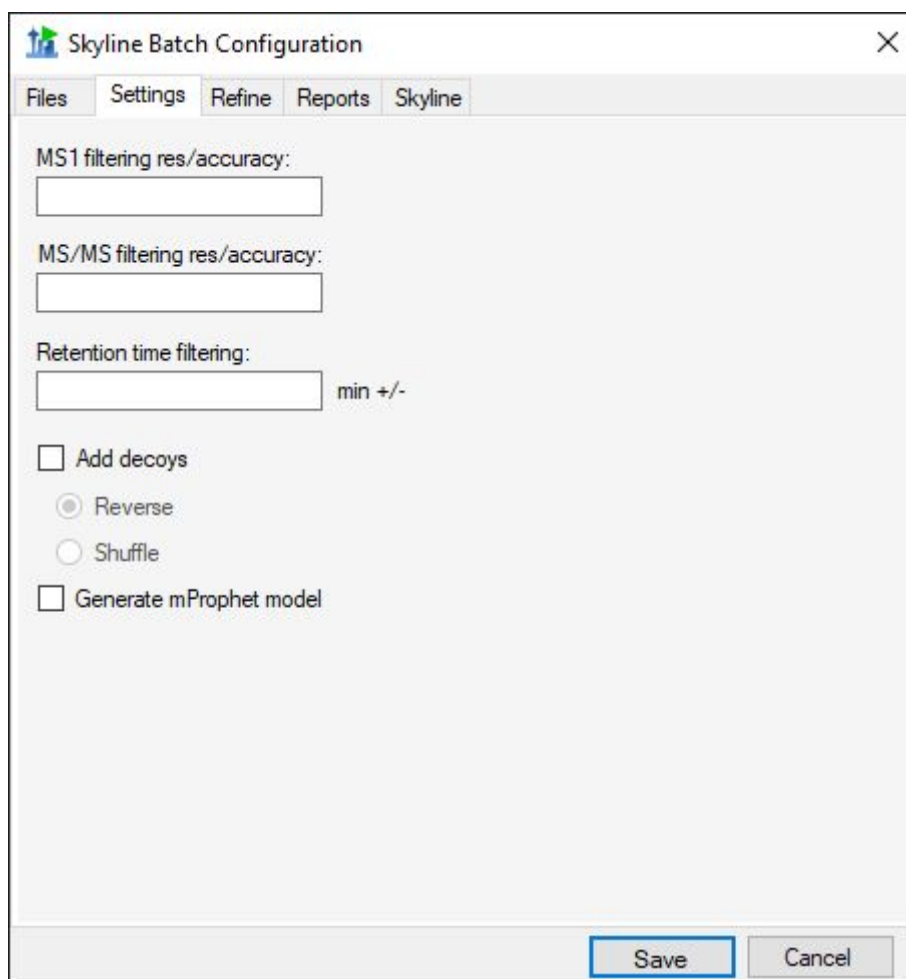All other download buttons bring up the similar **Download File From Panorama** dialog:

This will download one file at the specified location.

Remote source: The remote source for the file (see **Remote File Sources** section).
Relative path: The relative path to the file.

CONFIGURATION SETTINGS

In the **Settings** tab, various changes to the Skyline document can be specified.

MS1 filtering res/accuracy (optional): Resolving power of the precursor mass analyzer. If the mass analyzer is 'Centroided' the value is in PPM.
MS/MS filtering res/accuracy (optional): Resolving power of the product mass analyzer. If the mass analyzer is 'Centroided' the value is in PPM.
**Retention time filtering** (optional): The number of minutes on either side of the predicted time or MS/MS IDs, i.e. ± minutes.
Add decoys: Add decoys to the template document for reintegrate model generation with mProphet. Added decoys are shuffled or reversed based on selection.
Generate mProphet model: causes a new model to be created, using the mProphet algorithm with all available scores for the results found in the document. Existing manually integrated peaks will be overwritten with peaks chosen by the reintegration model.

CONFIGURATION REFINE

In the **Refine** tab, a new refined Skyline file can be created using the imported data.
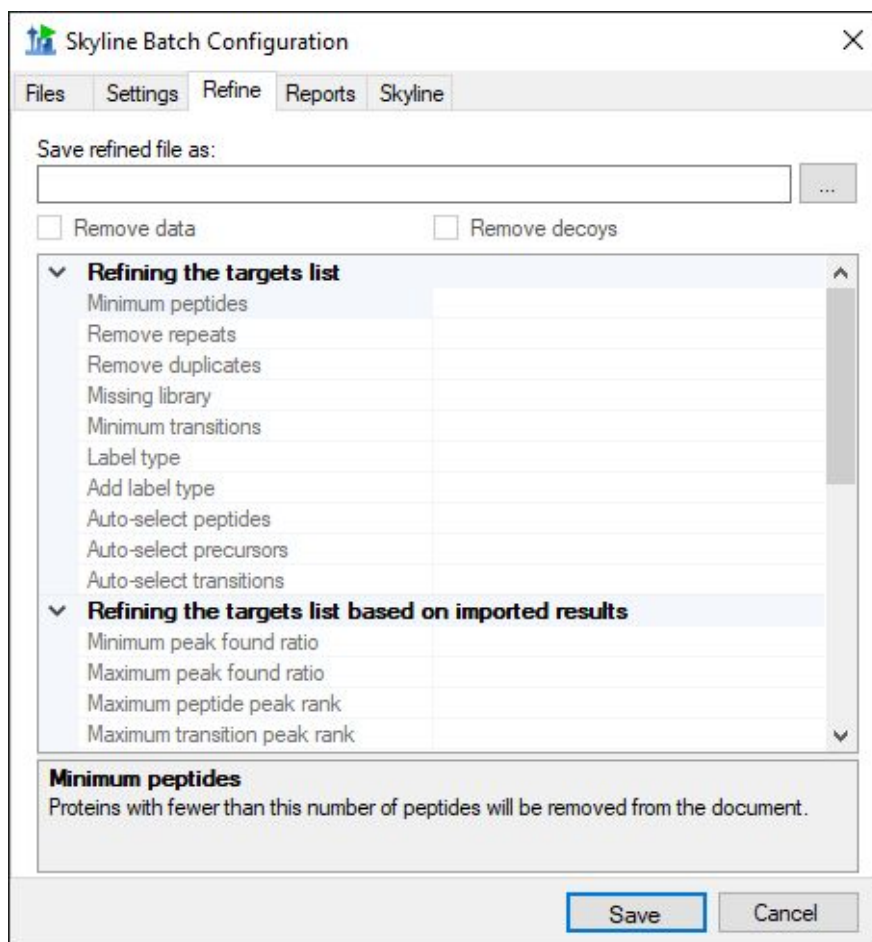
Save refined file as: The file path for the new refined file. This is required for any refinement to occur, and the rest of the **Refine** tab will be disabled until this field is filled.

> *Note: All the refine commands are performed on the refined file and will not change the results file in the analysis folder.

**Remove data**: If data should be removed from the refined file.
Remove decoys: If decoys should be removed from the refined file.

The descriptions for each of the commands in the grid can be found by clicking on the name of the command.
To enter a value for a command, click on the empty space next to its name. A list of options will appear, or you can type other input.

CONFIGURATION REPORTS

In the **Reports** tab, a list of reports can be created. If no reports are included for a configuration, Skyline Batch will not export reports or run R scripts for that configuration.
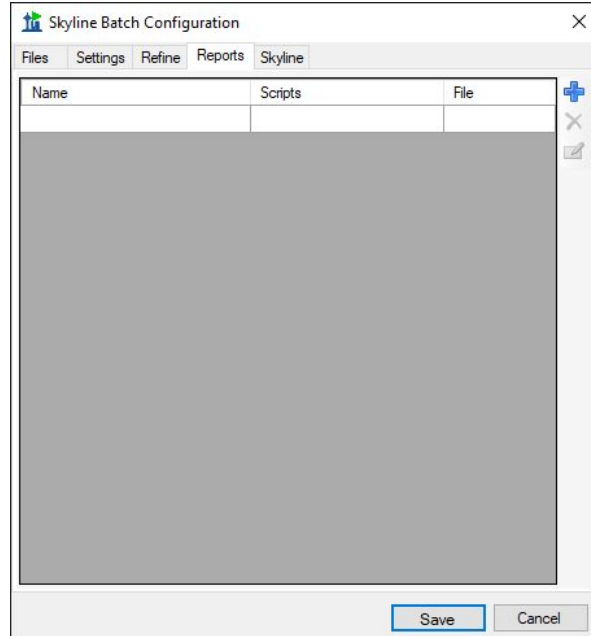
To add a report, click the blue **+** icon on the right and fill in the appropriate information.
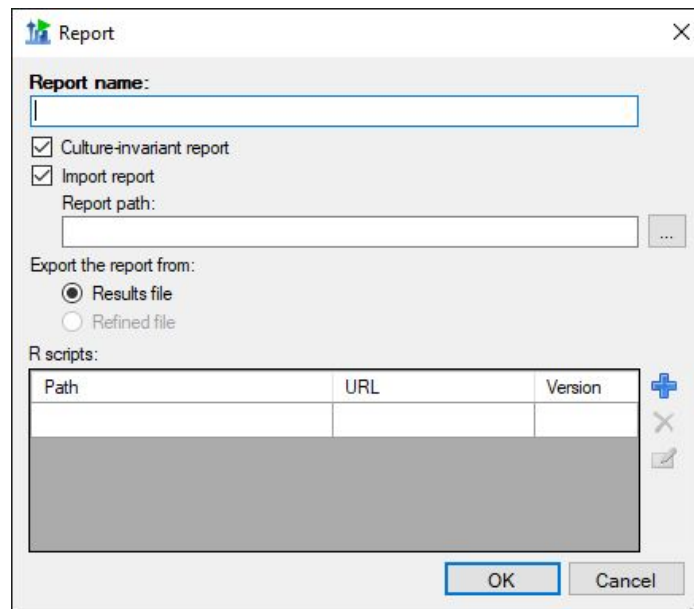
Report name: The name of the report to be exported. This MUST correspond to the name of the report when used in Skyline, which may differ from the file name. To find the name of a report:

If you are importing the report from a Skyline report file (.skyr), open the file in Notepad and look for: *name="*report name here"

If you are using a report already imported in your Skyline template document, open the document and go to View > Document Grid > Reports to see a list of all the report names.

Culture-Invariant report: If the report should use comma separated values with period decimal points. It is recommended you always use a Culture-Invariant report if you plan to run R scripts on the report.

Import report: If the report does not exist in your Skyline template document, and needs to be imported from a Skyline report file.

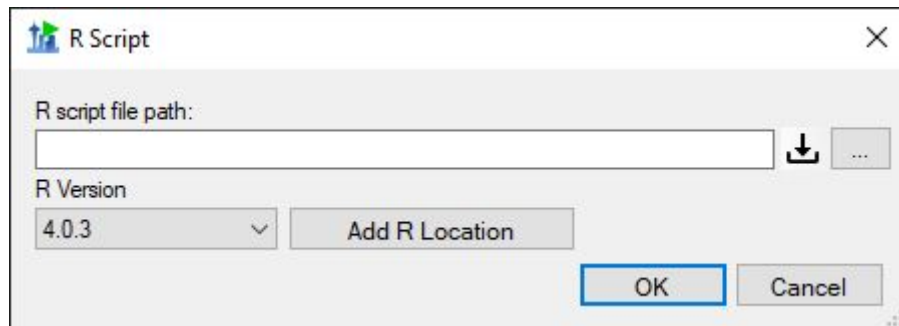**Report path**: The path to the Skyline report file.

Export the report from: Whether to export this report from the results file or the refined file. The refined file option will be grayed out if this configuration does not do refinement (ie: no file path was provided in the **Save refined file as** input on the **Refine** tab).

R scripts: The list of R scripts to run on this report after it has been exported to a csv file.

Add an R script by clicking on the blue **+** icon.

R Scripts

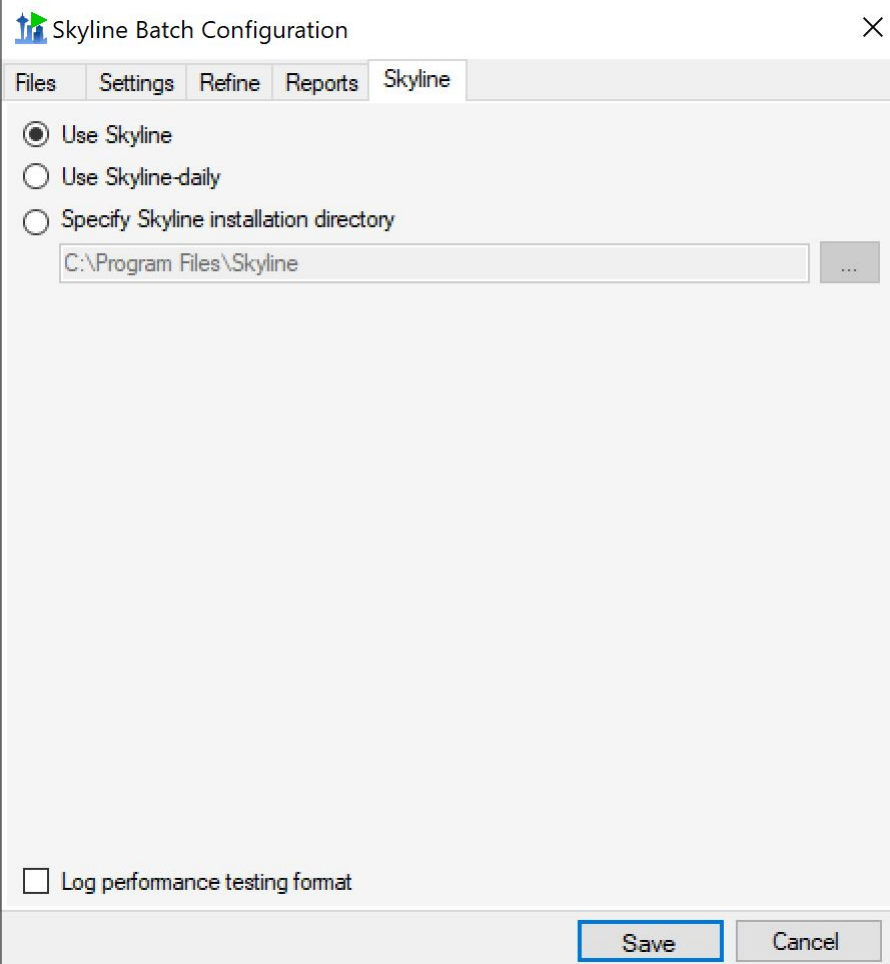The blue **+** icon brings up the **R Script** dialog:

R script file path: The path to the R script that will be run with this report as input. The R script will receive the report as a commandArg. You can read the data from the report and set the working directory to your analysis folder (recommended) with this R code:

```
reportPath <- commandArgs(trailingOnly = TRUE)[1] # get the path to the report
setwd(dirname(reportPath)) # set the working directory to your analysis folder
raw <- read.csv(reportPath) # read data from the report
```

R Version: The version of R that will be used to run this R script. The version of R can be changed in the drop-down list, or additional versions can be added to the list with the **Add R Location** button.

The download button indicates that this file can be downloaded from a server. Information about the **Download File From Panorama** dialog that appears when this button is clicked can be found on page 7.

SKYLINE SETTINGS

In the **Skyline** tab of the **Skyline Batch Configuration** form, you can specify which installation of Skyline this configuration will use when run. If you do not have an administrative or web-based installation of Skyline or Skyline-daily, the corresponding option will be grayed out. You can always specify a path to a Skyline (or Skyline-daily) installation somewhere else on your computer.

Log performance testing format: This controls whether or not the log for this configuration will be generated in a performance testing format. This format uses 24 hour time, is tab-spaced, and adds memory stamps to the output on the log tab. Unless you plan to analyze memory usage of the Skyline operations performed by the configuration, it is reccommended you leave this unchecked.

*Note: The Skyline tab does not appear in administrative installations of Skyline Batch. These installations use the version of Skyline they were installed with. If you have an administrative installation and want to change the Skyline Settings in this tab, download the web-based version of Skyline Batch here:
https://skyline.gs.washington.edu/software/SkylineBatch/index.html

SHARING CONFIGURATIONS

To save some or all of your existing configurations to a Skyline Batch configuration file (.bcfg), click the **Share…** button in the main window.
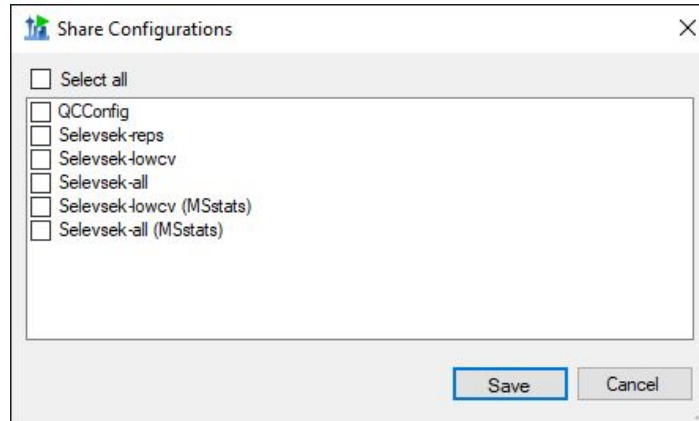

In the Share Configurations Form, you can select configurations you wish to save. Then click **Save**. Skyline Batch will then prompt you for a file location to save the

*Note: Skyline Batch can do better file location on new computers if the configuration is saved close to its files. This will result in fewer invalid configurations that need manual fixing. If you plan to share configurations across computers, it is recommended you save the configuration file in the same folder as one of the Skyline template files.
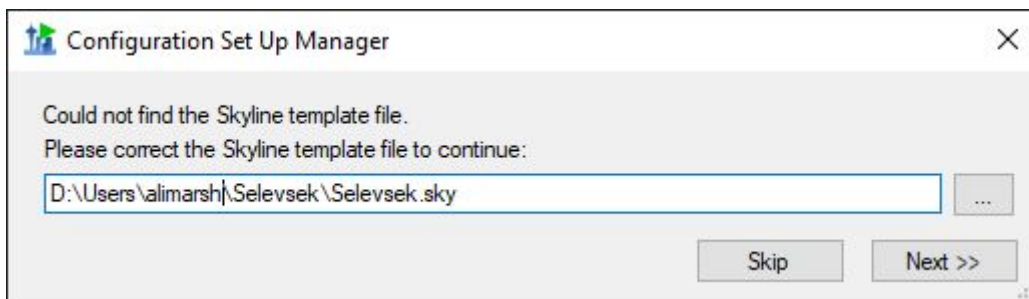
configurations.

The configurations in the resulting file can be imported using the **Import…** button, or by double-clicking on a Skyline Batch configuration file in File Explorer.
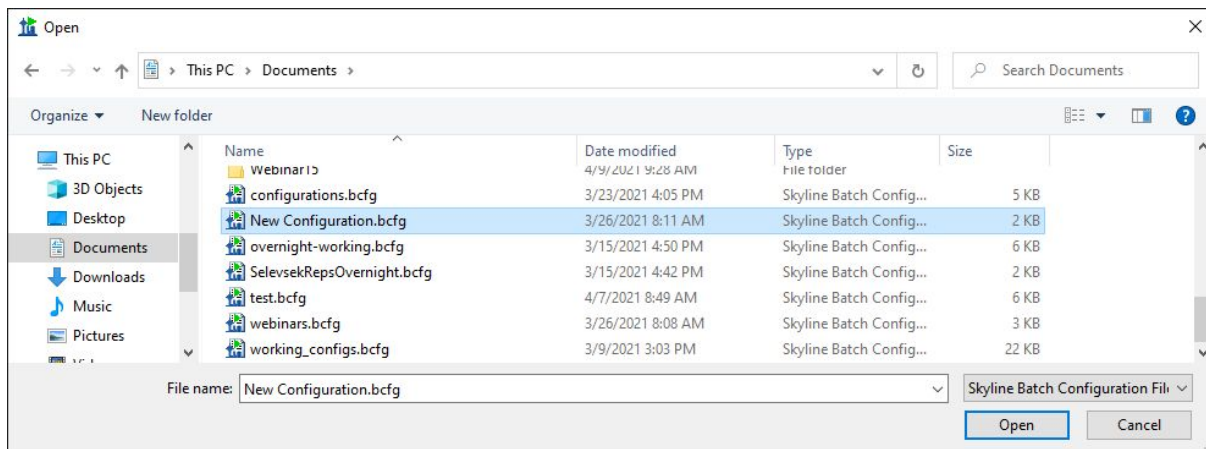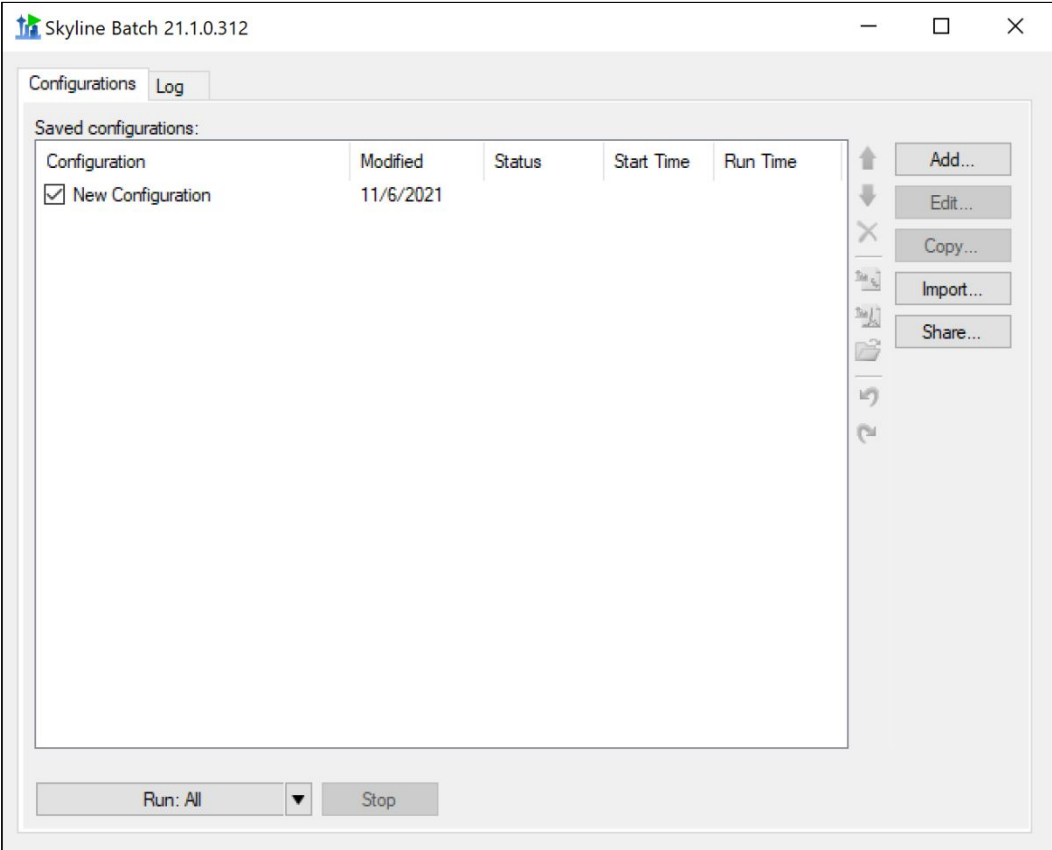
INVALID CONFIGURATIONS
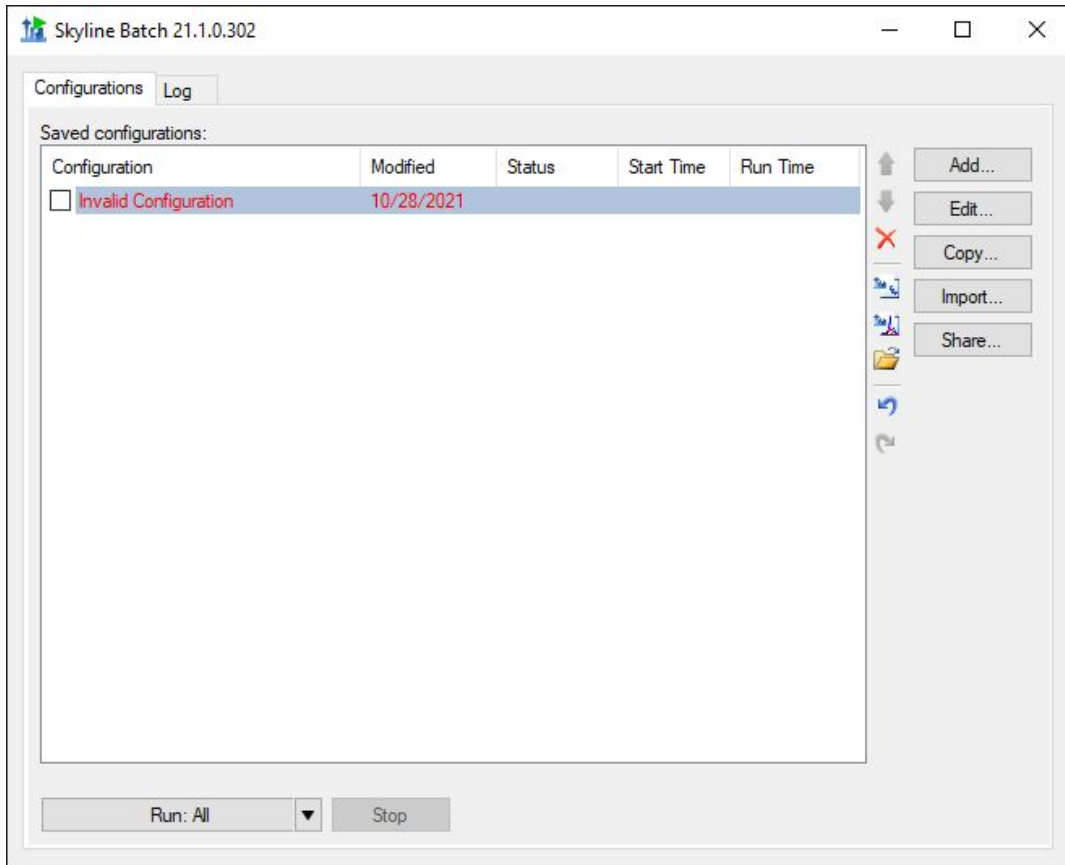
Invalid configurations may occur if files in the configuration do not exist, the configuration downloads files and could not reach the server, the configuration uses a version of R not found on your computer, or the Skyline version was not found. They will appear in red, and cannot be run until they have been corrected. When you **Edit** an invalid configuration, the **Configuration Set Up Manager** will guide you through the process of fixing it.
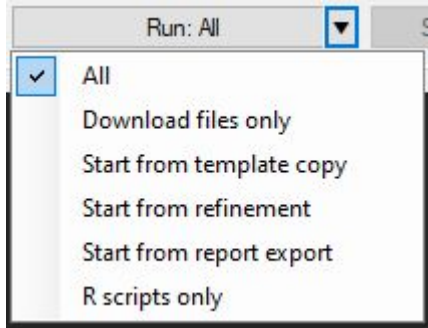
RUNNING CONFIGURATIONS

To run a configuration, check the checkbox next to it, then click the **Run All Steps** button.

The **Run All Steps** button will run every checked configuration in the list from top to bottom. The downward-facing triangle on the righthand side of the **Run All Steps** button enables the user to start from a later step when running the configurations.

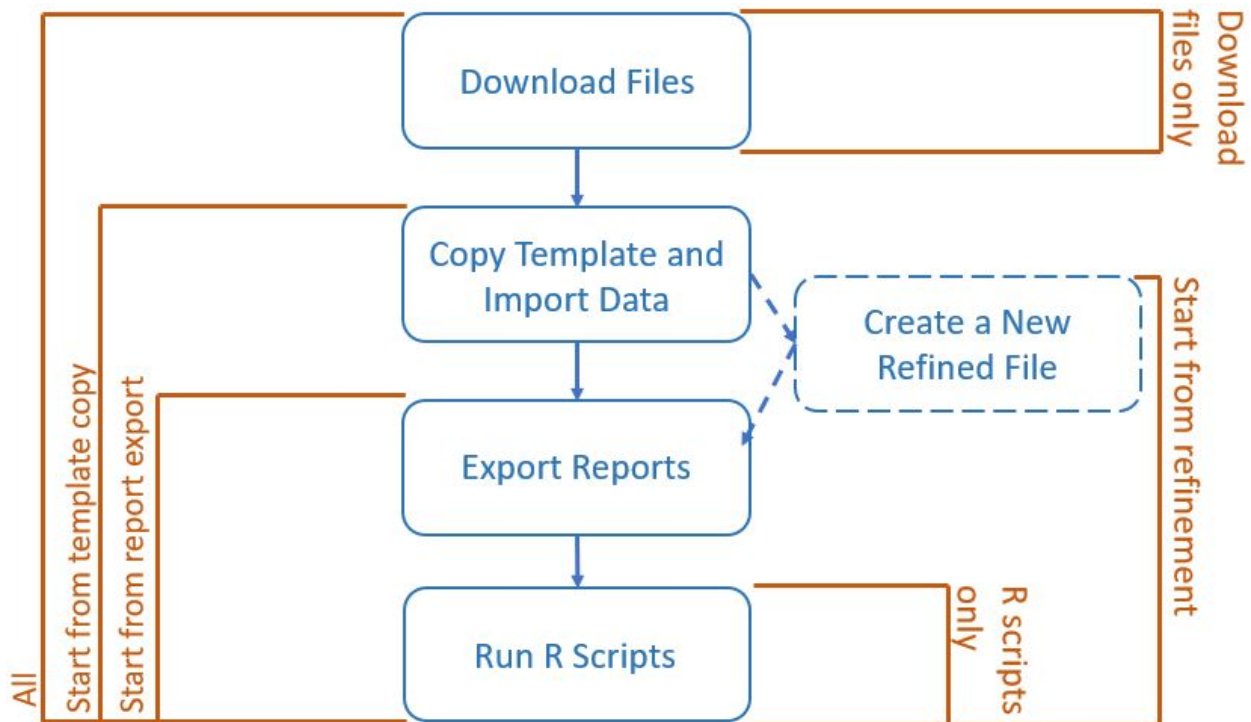The **Run from step** options depend on the configurations checked:

At least one configuration with a **refined output file** is checked:

No configurations with a **refined output file** are checked:
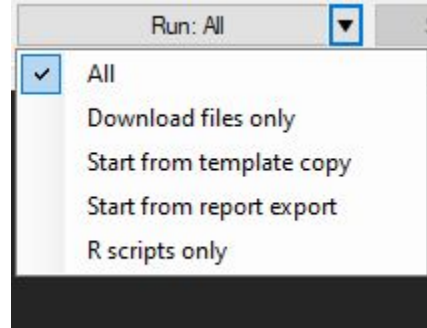
Each option runs the following section of the Skyline Batch workflow:

Keep in mind that a configuration without missing files, refinement, reports to export, or R scripts will skip the corresponding steps.

When the **Run All Steps** button or a **Run from step** option is clicked, the program switches to the **Log** tab to display the progress of the current run. Returning to the **Configurations** tab, the status of all configurations is displayed, and will show if they are Running, Waiting, Finished, Canceled, or had an Error. Stop the configurations from running by clicking the **Stop** button. If any configurations are running when Skyline Batch is closed, they will be canceled.

LOGGING

This is the view of the **Log** tab after the configurations have started running.
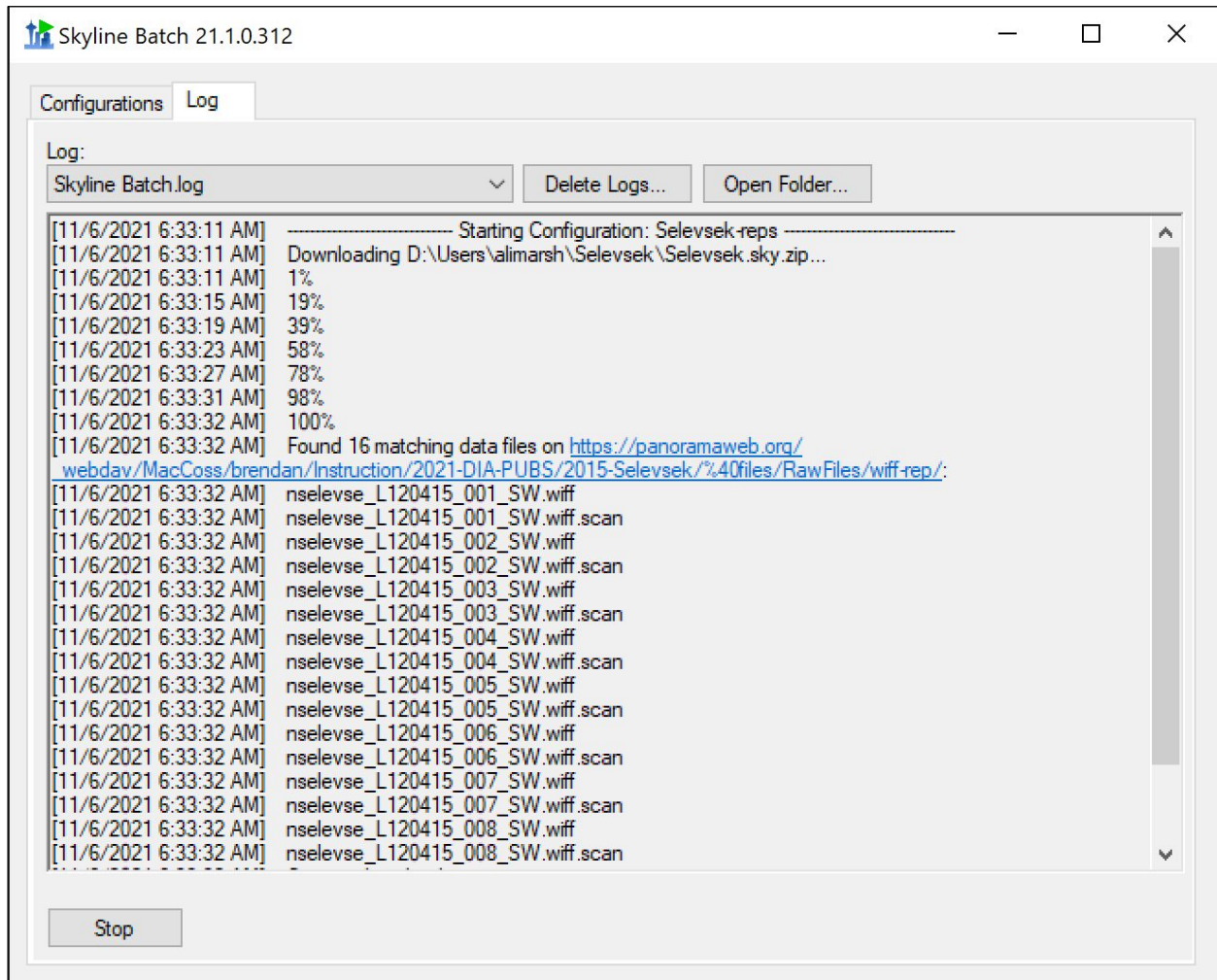
The log will display all output from running the configurations, including which configuration is currently running, what step it is on, and the progress of that step.
Logs of previous runs can be viewed by clicking the **Log** dropdown menu. Previous logs can be managed with the **Delete Logs** button and the folder containing the saved logs can be opened with the **Open Folder** button.

ACCESSING RESULTS

A configuration has finished running when its status no longer says "Running" and a Run Time is shown.
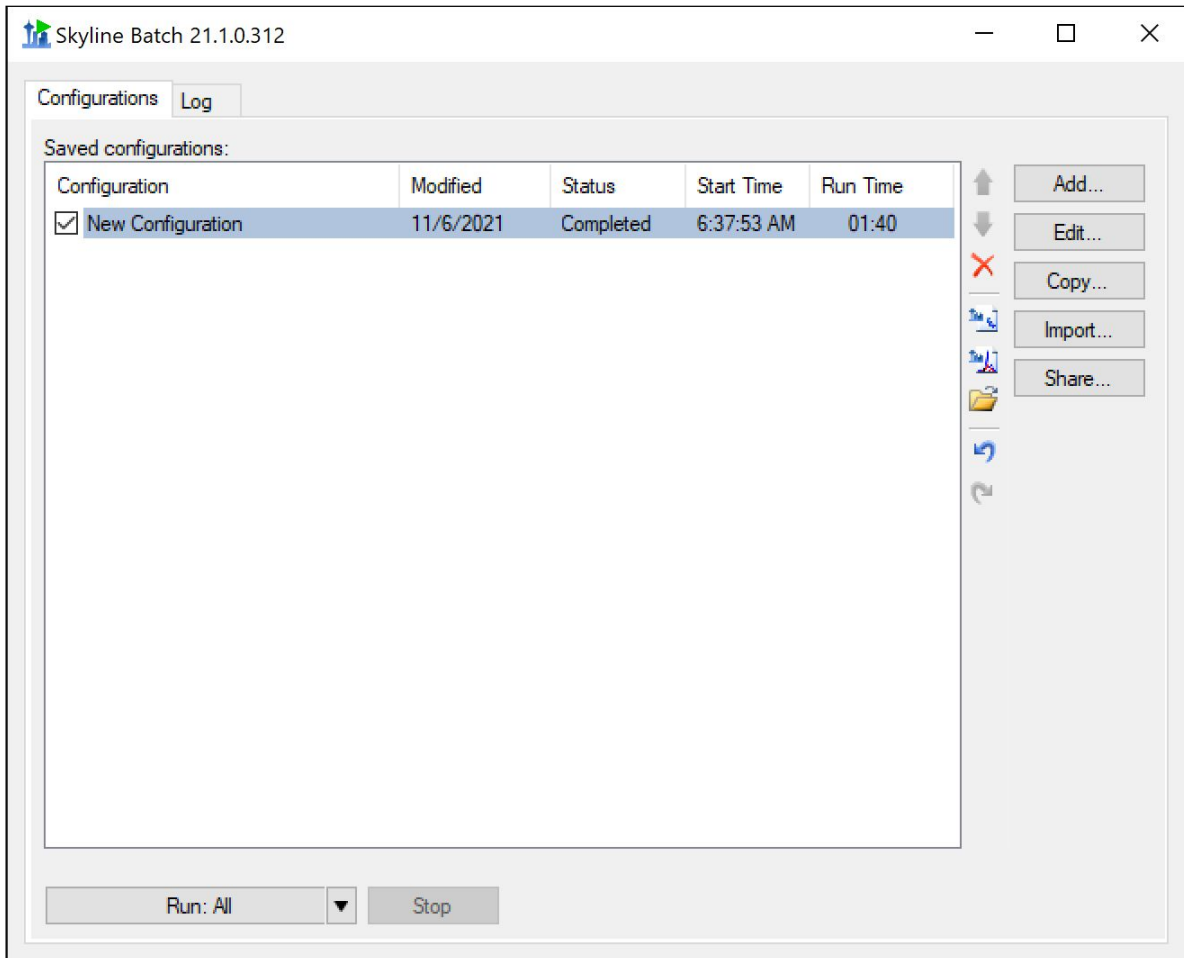
To view the results and associated files, select the configuration and click one of these icons:

 Opens the Skyline template document

 Opens the Skyline results document

 Opens the analysis folder

Supplementary Note 5. A basic tutorial to start using Skyline Batch with your own files.

This tutorial covers how to use Skyline Batch to create and run a basic configuration using local files. This is meant to be used with one's own files. If you are looking for a more concrete example with the files provided for you, the Skyline Batch instructions provided with Webinars 14, 15 or 20 may be more suitable (Start Page: /home/software/Skyline/events/2015 Webinars).

Creating a Configuration

- Download Skyline from https://skyline.ms/skyline.url
- Download Skyline Batch from https://skyline.ms/batch.url
- Open Skyline Batch using the start menu.

Skyline Batch looks for Skyline when you open it the first time. If it does not find it, you will see a form to specify the path to the Skyline directory. Do that.
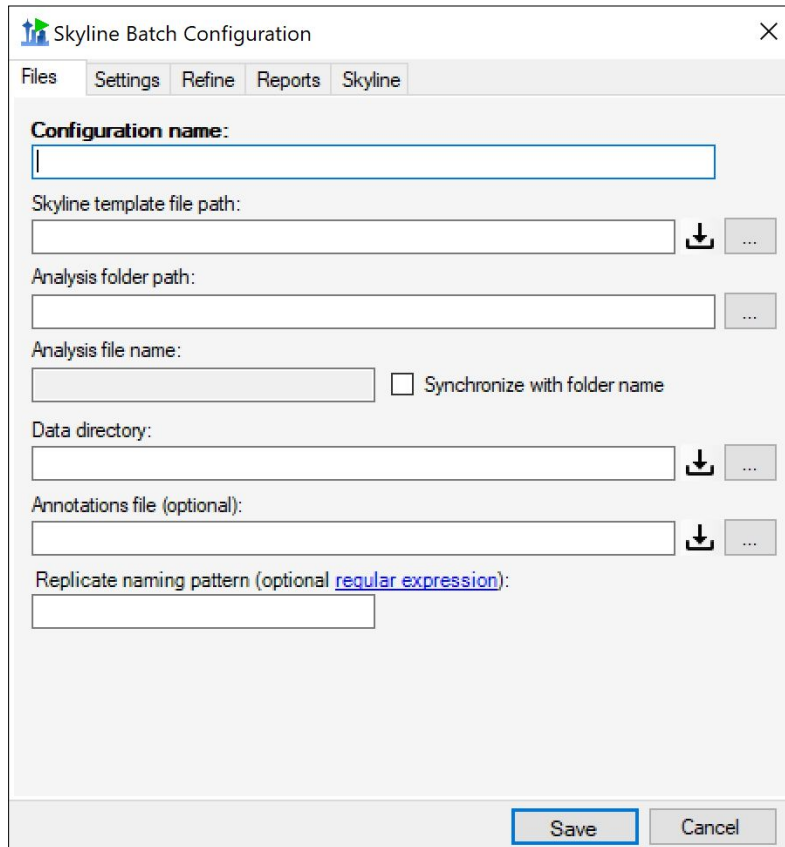
You will see the main window, which looks like this:
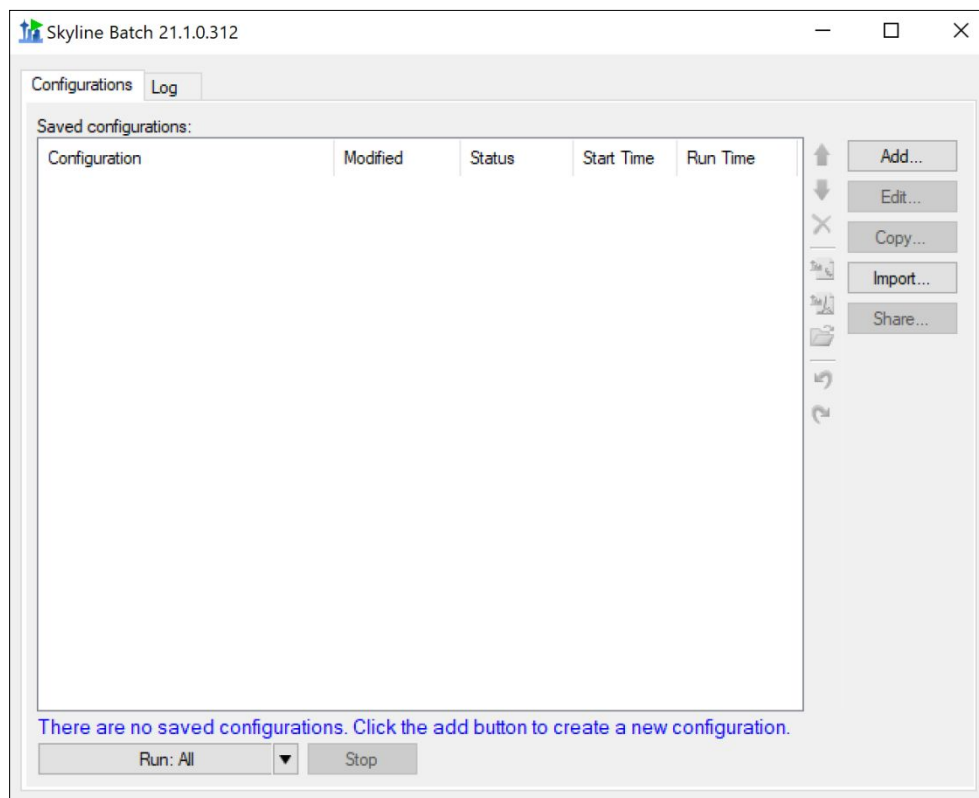
- Click the **Add** button.

This brings up the **Skyline Batch Configuration** dialog.

- Click on the **Files** tab.

Configuration name

This can be the name of the project or study for which data is being acquired. The configuration name must be unique.

- Select the **Configuration name** textbox and type your chosen name.

Skyline template file path:

This is the path to the template Skyline document that will be copied to the analysis folder and used to import data into.

- Click the "**…**" button next to **Skyline template file path**.
- Select your Skyline template file.

Analysis folder path:

The path to the local folder in which the copied template document and exported reports will be stored.

- Click the "**…**" button next to **Analysis folder path**.
- Select the folder you would like to use.

Analysis file name:

The name of the generated Skyline results file.

- Toggle the **Synchronize with folder name** checkbox to decide whether or not you want the file name to be based on the template file name or the analysis folder name.

Data directory:

The folder containing data to import into the copied template document. This folder should never contain anything aside from the data files you wish to import.

- Click the "**…**" button next to **Data folder path**.
- Select the folder you would like to use.

Annotations file (optional):

An optional CSV file containing the annotations for the imported data. Can be left blank.

- Click the (**…**) button next to **Annotations file** if you want to add one.
- Select the annotations file you would like to use.

Replicate naming pattern:

An optional regular expression from which the first group will be used to name replicates when importing. If a regular expression is used, it must match the names of the data files or it will cause an error when the configuration is run. Can be left blank.

- In the **Replicate naming pattern** textbox, type your regular expression if you want to include one.

The completed **Files** may look something like this:
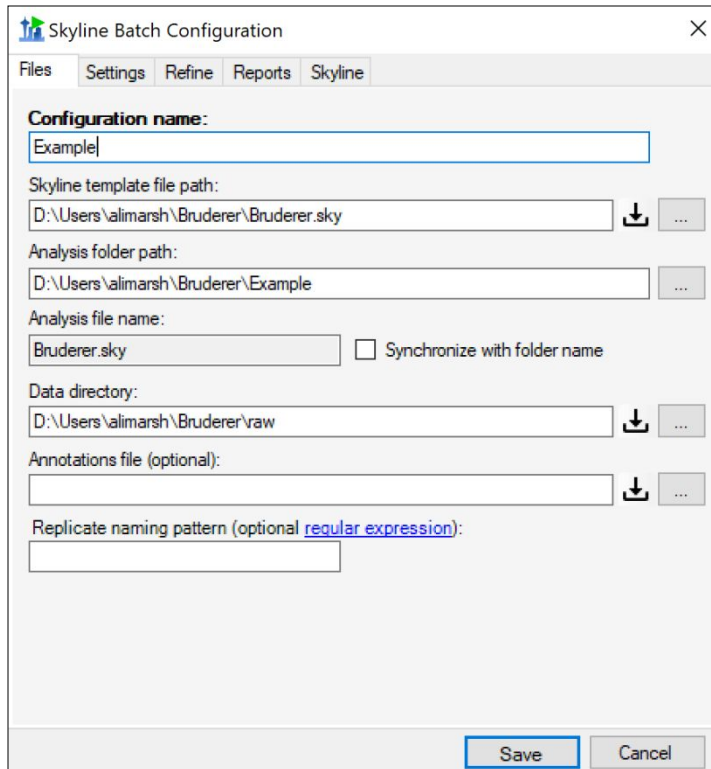
If this configuration is run as is, it will copy the template file to the analysis folder and import data into it. The rest of this section covers how to export reports and run R scripts with this configuration. If you don't want to export reports and run R scripts, skip to the "Running the configuration" section.

- Click on the **Reports** tab.
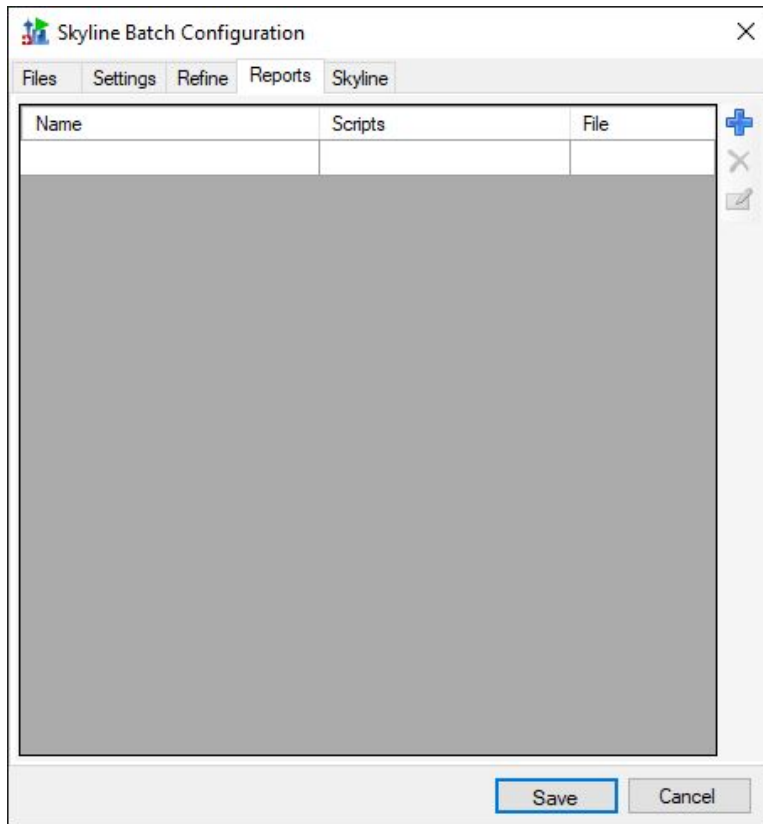
- Click the Add icon (➕) to add a new report.
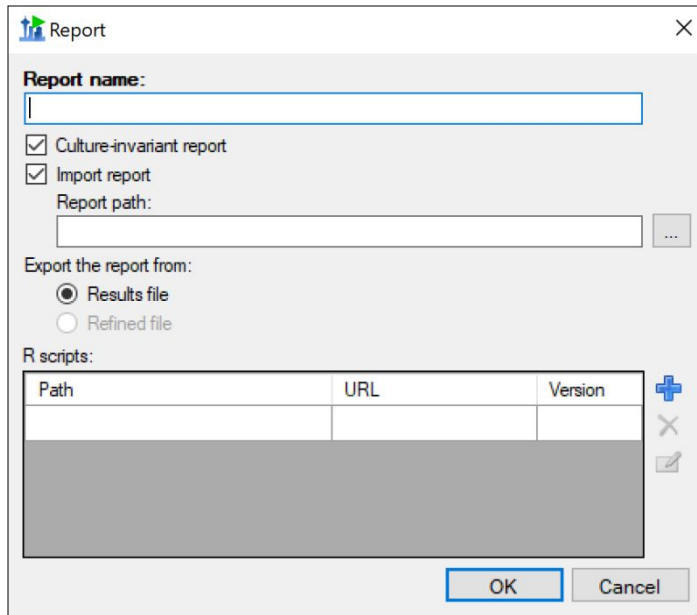
Report name:

The name of the report to be exported. This MUST correspond to the name of the report when used in Skyline, which may differ from the file name. To find the name of a report:

> If you are importing the report from a Skyline report file (.skyr), open the file in Notepad and look for: name="report name here"

> If you are using a report already imported in your Skyline template document, open the document and go to View > Document Grid > Reports to see a list of all the report names.

- Select the **Report name** textbox and the name of the report.

Import report:

This must be checked and a path to a Skyline Report file (.skyr) must be supplied if the report is not already imported into your Skyline template document. To check if the report has already been imported, open your Skyline template document and go to Settings > Document Settings. Select the Reports tab.

> If there is a checkbox next to the report name, it has been imported and you should:

- Uncheck the **Import report** checkbox

> If there is **NOT** a checkbox next to the report name, it has **NOT** been imported and you should:

- Click the (**…**) button next to **Report path**.
- Select the Skyline Report file you would like to use.

- Click the Add icon (➕) to select an R script if you want to include one.
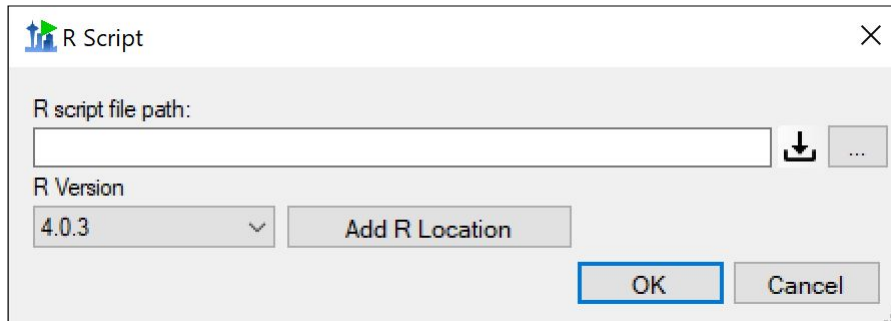
R script file path:

The path to the R script that will be run with this report as input. The R script will receive the report as a commandArg. You can read the data from the report and set the working directory to your analysis folder (recommended) with this code:

```
reportPath <- commandArgs(trailingOnly = TRUE)[1]
rootPath = dirname(reportPath) # in report dir
raw <- read.csv(reportPath)
setwd(rootPath)
```

When writing R scripts for Skyline Batch, it is also recommended you store all output in the working directory. Additionally, you should check for and install any necessary packages at the beginning of the script to increase the ability to share it between computers that have different packages installed.
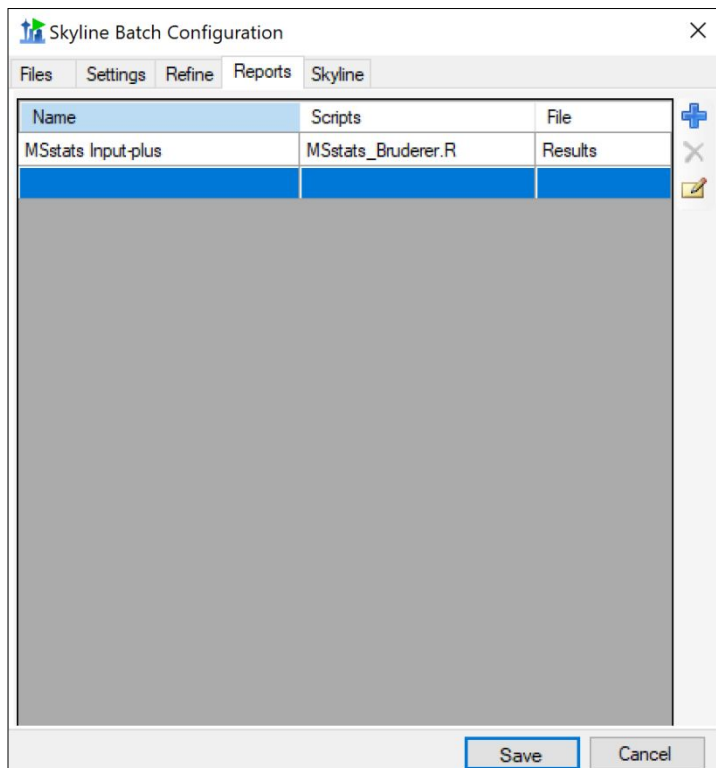
- Click the (**…**) button next to **R script file path**.
- Select the R script you would like to use.
- Select the **R version** you would like to use to run this script from the drop-down list.
- Click **OK** to add the R script.

The **Report** dialog may now look like this, with one R script in the list:
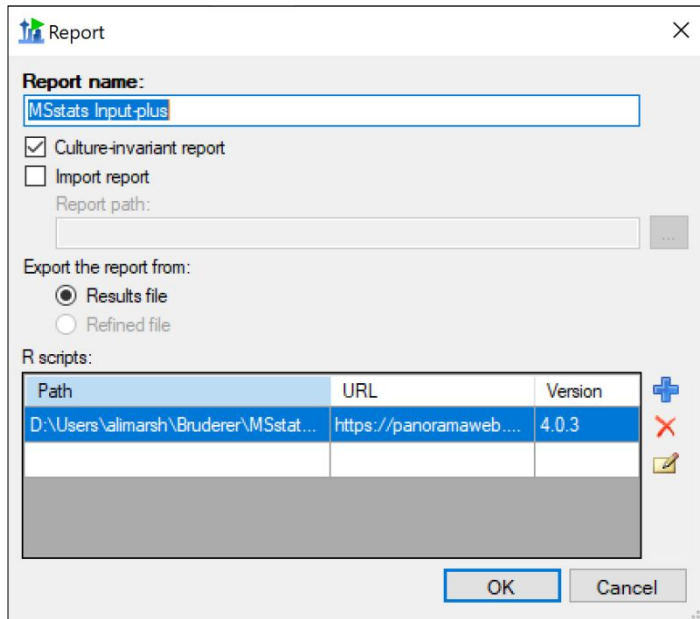
- Click **OK**.

The **Reports** tab may look like this, with one report in the list:
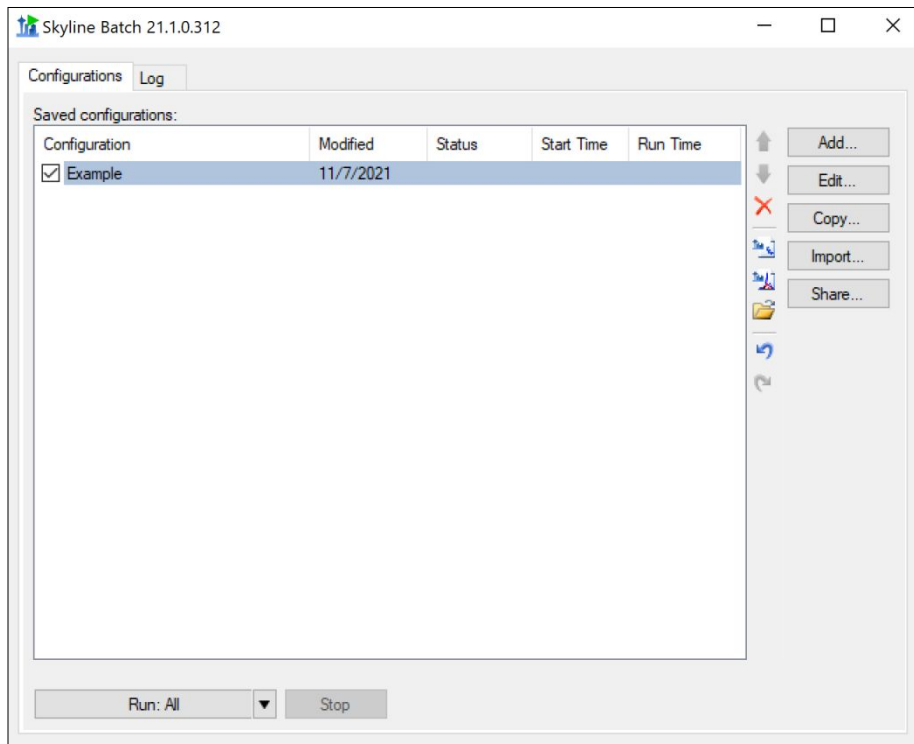
Running the Configuration

- Click **Save** to save the configuration.

If there are any errors in the configuration, a warning will appear that will prevent you from saving. Fix the information causing the error and try to save again.

The main window now has your configuration in the list:

- Click the **Run: All** button.

Skyline Batch brings you to the **Log** tab to show the output of running the configuration. When the configuration has finished running, it will display a **Completed** status in the main window.

View the output files by clicking on the configuration to select it, and using the following buttons:

- Opens the Skyline template document
- Opens the Skyline results document
- Opens the analysis folder

If you want to learn how to use other features in Skyline Batch, please visit the up to date documentation at:

SkylineBatchDocumentation.pdf