The following is a description of all the files you have at hand.

1. **X_M_train.npy** and **X_BC_train.npy**: two **training** sets of 13800 RGB images of size 250x250x3 collected from two different locations on campus. The images are panoramic 360°.
2. **{X_M_val.npy** and **X_BC_val.npy}** and **{X_M_test.npy** and **X_BC_test.npy}**: **validation** and **testing** sets of **2957** and **2958** RGB images respectively size **250x250x3** collected from two different locations on campus. The images are panoramic 360°.
3. **common_time_train.npy**: time of the collected training images. So, the first element of the array **common_time_train.npy** corresponds to when the first image of **X_M_train** and of **X_BC_train** were collected.
   Similarly for **common_time_val.npy** and **common_time_test.npy**
4. **ground_truth_train.npy**: measured GHI value at time of image collection. Example: the first element of this array is the GHI value collected at the first element in the array **common_time_train.npy**.
   Similarly for **ground_truth_val.npy** and **ground_test.npy**
5. **labels_train.npy**: GHI value 2 hours in advance from the time of image collection. In general, it will be a forward version (+2 hours) of the **ground_truth_train**. However, it won't always be the case as we stop collecting images at night so you don't have the **ground_truth** values at night and hence you won't be able to get the GHI value 2 hours in advance. That is why it is provided to you. Don't try to create these labels from the ground_truth data.
   So, the first element in **labels_train.npy** is the GHI value that corresponds to the first element in common_time_train + hours, i.e.
6. **clear_sky_val.csv (train, val and test):** This the value of GHI if there were no clouds in the sky. This is a physical model that can be computed for years in advance.
7. **meteo_data.csv (train, val and test)**: Many meteorological variables. Please check meteo_data_documentation.docx