5
Solar Radiation Gap Filling Approaches

5.1 INTRODUCTION

For identifying cloud condition and measurement error present in the available radiation dataset, different visualization techniques and quality control guidelines are suggested in the previous chapter. After comparing these, one can identify the approximate reason causing error in each solar radiation component. The radiation component and its possible interval which needs data correction is identified based on the source of occurrence. Primarily the data which failed to get stored in the database (due to any reason) are identified during the data step analysis. Here data gap filling is compulsory; as it may have occurred due to some human maintenance schedule, hardware issues or in reality no data is present at that location (due to any predictable condition). Next case is data identified as erroneous, after application of quality control guidelines. Nowadays, all stations measure radiation at a high-frequency rate (1-sec to 1-min), so missing instances of data are possible. However, the interval for which data is found erroneous, a proper investigation of climate conditions and instrument response needs to be studied and the final decision is taken regarding its corrections.

Once the missing interval to be corrected is identified, according to the identification of gaps and their frequency, these gaps can be filled either manually or semi-automatically. In manual gap filling process, each gap is filled by using location averages, interpolations or standard equations. In addition, by using this manual approach, a semi-automatic flowchart is suggested by Geuder *et al.,*(2015) and Kumar *et al.,* (2014), which fills the missing values, according to the data missing interval in the database. To automate this process, high-quality data servers and processing machines are used, for storing and processing the measured data recorded from the instruments. In addition comparison of currently used gap filling approaches is discussed below, with their gap filling guidelines for data of various frequencies. For validation of these, from available radiation databases, different cloud condition days are selected and gap filling is done using various approaches. Selection computationally less intensive approaches and location-specific modification are suggested below.

Selected radiation database is of 10-minute interval; hence most relevant gap filling techniques are to be identified for this. Now after performing a detailed study on all possible ways of filling the missing gaps from the literature available, some approaches are finalized. These approaches are selected from the research work carried out for different locations around the world, but when applied outside these locations they may not perform as well. Hence fitting the gaps using transmittance ratio [Maxwell *et al.,* 1993], meteorological values [Geuder *et al.,* 2015], curve fitting [Tejera *et al.,* 2015] and comparable satellite data [Kumar *et al.,* 2014] are the only identified approaches which are found valid. In this chapter, each approach will be discussed individually with their available guidelines.

The process of gap filling, involves combination of available relations, standard equations and curve fitting methods. But for each approach, its gap prediction conditions and computation required are the key deciding factors for its selection. Complex physics-based computational procedures may be accurate but require high investment (additional databases and computational cost). In case of high-frequency data measured, some identified cases of data missing may not affect the overall site assessment. So researchers usually take the simplest route of either using "linear interpolations" or direct gap filling using "curve fitting" approaches. These approaches can easily fill the missing gaps if data missing duration is small. But if gap frequency is greater than 3 hours, then predicted data don't fit the actual measurement pattern. Also for different climatic conditions, all approaches may differ from each other.

5.2 IDENTIFICATION OF GAPS IN MEASURED DATABASE

During the measurement of different radiation components, the measurement error can arise in different intervals of any radiation component. Any possibility of a combination of measurement errors for a given interval is possible. For example, only one component missing for the whole interval or for some interval, data for all components is erroneous. Hence for these missing components, individual gap (missing) identification is important. As for different data missing cases, the data filling approach can be different. Analyses of different types of gaps in radiation database are done according to Schwandt *et al.,* (2014). Types of gaps are identified as single-gap (out of 3 component any 1 component is not present); double gap (GHI is present but both DNI and DHI are missing); triple gap with (all 3 component missing) data missing for less than three intervals and triple gap with data missing greater than three intervals to up to 1 day (see Table 5.1). Now with this identification, relevant gap filling process needs to be selected.

Table 5.1: Identification of Various Gaps Present in Measured Radiation Database

Sr. No.	*Gaps*	*Explanation*
1.	Single Gap	Out of three radiation elements, anyone is missing
2.	Double Gap	DNI and DHI are missing and GHI present
3.	Triple Gap	All three element missing for 2-3 intervals
4.	More than Triple Gap	All three elements missing for greater than three intervals

Note: For the double gap-filling procedure, the GHI should always be present and gap filling is required for DHI and DNI components. However, if GHI is not present, then consider this missing case as triple gap filling. GHI radiation component theoretically consists of DHI and DNI measured values, according to the physical radiation relationship between radiation components. Hence, estimating either DHI or DNI component from GHI component is possible but if anyone of DHI or DNI is missing, then GHI can't be calculated.

5.3 TRADITIONAL SOLAR RADIATION GAP FILLING APPROACHES

Identification of data missing from the database for any location is discussed above. Now gap-filling approach should be used for filling these missing intervals in radiation database. Therefore, for missing gaps in databases with yearly, monthly and daily averages, vast amount of literature is available. These techniques range from the use of basic averaging to the latest machine learning algorithms [Zhang *et al.,* 2017]. These can be applied for filling data with different frequencies. But for sub-hourly values, no standard approach is available in the literature. However, before filling for missing values, data reliability is questioned. If the database is not reliable, then it's processing (gap filling) will also become erroneous. Some physical parameters are identified in Chapter 4 which affect the radiation at a given site. Without being aware of these component error contributions, one can't accurately predict the missing values from the available database. The case of dust present on the sensors is an example in this context.

The most common and well-established approach for gap filling is by calculating mean values, covariance matrices analysis and data imputation (statistical models, skewness and kurtosis) techniques [Schneider, 2001; Pappas, 2014; Ludstrom, 2012; Meyer, 2008; Hay, 1993]. Moreover, historical radiation database if available can be used to fill the gaps. Calculating radiation values using empirical relations, based on climate conditions is also an option. There are different computer simulation packages [Kandasamy *et al.,* 2013] for this task. If nearby location meteorological station databases are available then missed data is filled using these. Finally, after gap filling, the comparison is done between raw and modified databases [Ridley and Boland, 2008]. A summary of gap filling approaches from the literature are listed below (these relations are valid for all frequency datasets).

5.3.1 Data Filling Using Averaging and Interpolation Function

This is a simple approach available to fill the missing values in the radiation database. According to the types of gaps identified above, for single data missing case, direct averaging is done between two available values. Whereas for greater than that, one has to find the difference between two available values, and then accordingly in between gaps are filled. In addition, this approach is available in a number of mathematical software packages and by doing interpolation and extrapolation data filling is done. Here available linear and non-linear functions can be used [Modak, 2010] and the majority of researchers prefer this method for filling any frequency data gaps.

5.3.2 Data Filling Using Various Available Clear sky and overcrowded Models

Another approach used for filling the missing values of hourly and daily averaged values is by using standard relations, like clear sky equations (discussed in chapter 2) and different cloud condition equations (for ex. Linke turbidity equation, for diffuse component calculation). In these models, input data required is historically available values, climatic derived parameter equation and location specific relations for data filling.

5.3.3 Manual correction of data

If gaps cannot get filled by any of the techniques given above then manual correction approach can be used. This condition occurs when data gap is large and no supporting database is available for its approximate filling. Measurement error occurring due to instrument regular maintenance cycle, instrument tracking offset selection, etc. are some cases, in which one can't predict the possible reason for the occurrence of an error, but still correction is required. Hence according to the operator's experience, these values are corrected manually.

Using above gap filling approaches automatic and semi-automatic approaches are suggested in the literature. These approaches are reviewed in [Kumar *et al.*, 2014, Geuder *et al.,* 2015; Maxwell *et al.,* 1993]. They provide a list of possible techniques, by which for any solar radiation data set, data gap filling can be done.

5.4 SUGGESTED GAP FILLING APPROACHES FOR HIGH-FREQUENCY DATASET

The database selected for analysis is having 10-minutes interval radiation database. From the available literature, gap filling approaches found applicable for this measurement frequency are identified. Where for each type of gap (single, double and triple) identified at different intervals, unique gap filling guidelines are suggested as given below.

5.4.1 Gap Filling Using Meteorological Parameters

In some cases one has an acceptable sub-interval meteorological data (ex. solar zenith angle) available, then for that missing interval, radiation values can be calculated. Also by using relations derived from relative humidity, atmospheric temperature, cloud cover and local precipitation values, radiation can be estimated. Moreover, using transmittance models like diffuse fractions [Reindl *et al.,* 1990] and clearness index [Batlles *et al.,* 2000] for any location, specific gap prediction relations can be derived (correlations between selected parameters and radiation component).

Spokas *et al.*, (2006) suggested some relations, which are used for calculation of hourly averaged radiation component values using meteorological parameters (Relative Humidity, Barometric Pressure and Solar Zenith Angle) (see Eq.(5.1 to 5.5)). The controlling parameters of this equation are uniformly averaged ground measured relative humidity values (see Table (5.1), which show a correlation between relative humidity and atmospheric transmittance (derived from US location database). For identifying remaining required parameter values, they are either taken from the given database or calculated theoretically from standard equations [Sukhatme *et al.*, 2008]. Here Eq. (5.5) shows direct "physical radiation relationship" between three measured radiation components.

$$\cos(z) = \sin(\phi)\sin(\delta) + \cos(\phi)\cos(\delta)\cos(\omega) \tag{5.1}$$

$$m = \mathrm{Pr}s/(101.3(\cos(z))) \tag{5.2}$$

$$DNI = 1365\,(\tau)^{m} \tag{5.3}$$

$$DHI = 0.30(1-(\tau)^{m})1365\,(\cos(z)) \tag{5.4}$$

$$GHI = DHI + DNI(\cos(z)) \tag{5.5}$$

where,
ϕ, δ, ω, z= latitude, declination, hour and zenith angle
τ, m = Atmospheric transmittance and optical air-mass
Prs = Barometric pressure

Another available approach uses ratios of meteorological parameters (location specific maximum and minimum temperature and station elevation) with a local coefficient and standard solar Extra-terrestrial coefficient [Hargreaves *et al.,* 1982; Annandale *et. al.,* 2002]. But due to the use of daily average values, the results obtained from this model are found inferior to other sub-hourly suggested approaches. For gap filling of a single component, the physical radiation relationship is directly used. Whereas for further gaps, equations mention above are used.

Table 5.2: Correlation between Measured Relative Humidity with Atmospheric Transmittance

Sr. No.	*Relative Humidity (%)*	*Atmospheric Transmittance(τ)*
1	RH≤40	0.69
2	40≤RH≤45	0.67
3	45≤RH≤55	0.57
4	55≤RH≤65	0.47
5	65≤RH≤75	0.41
6	75≤RH≤80	0.3
7	80>RH	0.2

5.4.2 Gap Filling using Approach Given by RMIB

This approach provides the most advanced data quality and standard gap-filling guidelines, applicable for European regions (Belgium, Spain, Germany and France). Here suggested gap-filling procedure is a mixture of components of inter-correlated relations, and input data is calculated from given location historical averages and standard relations. Now for determining each solar radiation components (GHI, DHI & DNI), individual equations are provided here (see Eq. (5.6 to 5.9)). The models suggested are developed for calculating daily or hourly averages, but they can also be applied for sub-hourly (i.e. 30-min and 10-min) dataset.

In this gap-filling approach, the missed components are calculated individually. Where for calculating direct radiation component at any specific interval, Erbs *et al.,* (1982) and Skartveit & Olseth, (1987) suggested an approach (if the GHI value is available). But if global component is not present for that interval, then direct radiation component is calculated, using Extra-terrestrial values (I_E) with Maxwell, (1987) and Perez *et al.,* (1992) suggested model. Next for estimation of GHI component, Angstrom, (1924); Almorox *et al.,* (2004) and reverse Skartveit *et al.,* (1987) approach is used. Here the input data required for determination of gap values, by these models are the average sunshine duration and location-specific constants (calculated using approach given in Sukhatme *et al.,* (2008)). Using location specific plot of clearness index and beam transmittance (for different solar altitude angles), the global value is determined.

$$k \le k_o \; ; H_d / H_g = 1.0 \tag{5.6}$$

$$k_o \le k \le \beta k_o \; ; H_d / H_g = f(k) = 1 - (1 - d_1)(j\sqrt{K} + lK + (1 - j - l)K^2) \tag{5.7}$$

$$K = 0.5\left(1 + \sin\pi\left(\frac{k - k_o}{k_1 - k_o} - 0.5\right)\right) \tag{5.8}$$

$$k \ge \beta k_1 \; ; H_d / H_g = 1 - \beta k_1 \left(1 - f\left(\beta k_1\right)\right) / k \tag{5.9}$$

Here k_o, k_1, d_1, j & l are to be determined using least square analysis on historical dataset.

$$\frac{I_G}{I_E} = a + b\frac{SD}{SD_{\max}} \tag{5.10}$$

If no radiation data is available for any measured interval (data missing for greater than 3 intervals), then for calculation of global component Angstrom, (1924) (see Eq. (5.10)), and calculation of the ratio of sunshine hours is done using standard theoretical relation [Sukhatme *et al.,* 2008]. In addition, Journée and Bertrand, (2011) have suggested 8 model equations for the computation of the sunshine hours. Here these model equations are validated at various sites in Belgium, but not for any other location. Hence by using standard theoretical relations, with available location specific constants and clear-sky index, the missing data are filled. Finally, the most preferred sunshine model is selected for comparing this generated value.

To conclude, for the single gap-filling procedure, direct physical radiation relationship is used (see Eq. (5.5)). Whereas for all cases of double and triple gap-filling, the suggested relations given above in Eq. (5.6 to 5.9) are used for estimation of DNI, GHI and sunshine ratios. The remaining DHI value is calculated by using the "direct physical radiation relationship" equation.

5.4.3 Gap Filling Using Approach Given by GIZ (C-WET)

Gap-filling procedure, used by all ground-based (C-WET) SRRA measurement stations [Kumar *et al.*, 2014] is discussed here. Till date, 121 SRRA basic and advance stations are installed all over the India. These provide good quality radiation datasets. For maintaining database quality, guidelines for gap identification and their subsequent gap filling approaches are described by Schwandt *et al.,* (2014) are used.

Now for single gap filling problem, physical radiation relationship equation (Eq. (5.5)) is used, whereby using available two radiation components with solar zenith angle, the third missing component is calculated. Whereas in double gap-filling, guidelines given by "RMIB" is used. Hence diffuse component is calculated by using conditions suggested by Skartveit *et al.,* (1987), with the compulsory condition of GHI value presence for that interval. Moreover, for small gaps, the diffuse horizontal transmittance or clearness index is calculated for respective solar elevation angle and by using the interpolation technique, the missing component is calculated. Now as two solar irradiance components are present, by using physical radiation relationship, missing third radiation component is calculated.

For the triple-gap filling, where all three solar components are not present for less than 3 intervals. Transmittance ratio for GHI and DNI component is calculated for available values and by using linear interpolation, the missing value transmittance factor is calculated. Now from these ratios, GHI and DNI are extracted, and finally, by using physical radiation relationship equation, the DHI values are calculated. But if all three solar components are missing for greater than 3 intervals (up to 1 day), then the most optimal method is to take the missing data from the previous or near available day. As the solar radiation variation for near days is expected to be small, and no change is seen in solar elevation and zenith angle for that location.

5.4.4 Gap Filling Using Satellite Database

It is used for gap filling of any interval radiation database. The gap filling is done by comparing the same time-stamp satellite data with ground-based measured radiation values (with is having gaps) [Schumann *et al.,*2011; Schwandt *et al.,* 2014]. This gap-filling method is only applicable, when satellite data is uniformly measured for that location, with no missing values. Also, the selected satellite dataset should be spatially precise and have minimum 60-minute time resolution. Then it can be easily overlapped with ground measured data. Hence for data comparison and analysis, both time series radiation datasets of the same frequency are plotted together. Now the trend or pattern followed by satellite data is compared with ground database and accordingly missing values is filled.

For further understanding of the potential of this suggested gap-filling approach, one can consult Kumar et al., (2014). Here 9 SRRA radiation measurement stations located in different parts of India representing different climatic conditions are selected and artificial gaps are uniformly introduced in the database. Now using suggested gap-filling approach, the missing values are predicted. Comparing both measured and predicted database, each station's relative mean bias value is calculated for each interval and finally by taking the average over all 9 SRRA stations, the combined result is presented. For a gap filling interval having only single component missing, the gap filling guidelines predict missing value within "± 1" percent range. Whereas for double gap filling, due to presence of cloud conditions, the variability of measured radiation parameters is frequent and the relative mean bias is identified high (DNI is increased by 19% and DHI is decreased by 9%). Now for higher gaps, where all the three components are missing (but for a small duration), results produced for each component are decreased by 1-2%. At last, for greater than 3 interval duration gap filling, the results are in the range of 5-10% for different radiation components (where previous day data is used for gap filling).

5.4.5 Gap Filling Using Curve Fitting Approach

This approach is the most applicable method, for filling the gaps in any time-series dataset. Here for the available dataset which is having missing values, curve fitting is used for that interval. Now the curve fitted for that duration easily fills the missing values. In this process, both linear and non-linear approaches are applicable [Karim *et al.,* 2014]. Moreover, in an analysis available in the literature, uses of polynomial fitting (n= 1 to 9) and sinusoidal fitting (see Eq. (5.11 and 5.12)) are suggested for filling missing gaps on all cloud condition days. A quadratic polynomial is found to give good results with less computation required.

$$f(x) = p1(x)^9 + p2(x)^8 + p3(x)^7 \ldots\ldots + p9 \tag{5.11}$$

(Polynomial fitting)
n=1,2,3,4,5,6,7,8,9
where, p1,p2,p3……,p9= coefficients with 95% confidence bounds

$$f(x) = a1(\sin(b1(x) + c1) \tag{5.12}$$

(Sinusoidal fitting)
where, a1,b1,c1= coefficients with 95% confidence bounds

5.5 COMPARISION OF VARIOUS GAP FILLING APPROACHES

All gap filling approaches discussed in the previous section yield databases that are better than the raw database. The selection of the best one needs other considerations. Moreover, the cloud conditions, dust deposited on the sensors also reduces the gap prediction ability of these approaches. Hence to clearly understand each suggested gap-filling approach in detail, a comparison is done here (see Table 5.3).

5.6 DETERMINE GAP FILLING APPROACH FOR DIFFERENT CLOUD CONDITIONS

A study is done to carry out gap filling based on cloud condition. In this analysis, during data selection and preparation, some rainy days are selected from the provided dataset (having no missing intervals). These days are arranged based on type of cloud present (clear sky, lightly clouded, medium clouded and heavily clouded) during that day. The transmittance values are calculated for each interval, which is used to verify local climatic conditions and measurement error present in the database. In further analysis, suggested quality control guidelines are applied and measurement error present in each radiation component is flagged accordingly. Once the data passes all tests, artificially some gaps of a different frequency are introduced in this database. This modified database is now gap filled using approaches discussed above. Here the analysis is done on the basis of different data classification components (different gap types, cloud conditions and gap filling approaches). Based on this analysis, a guideline is given for the selection of gap filling techniques at different gaps present for various cloud conditions.

5.6.1 Sample Data Preparation for Gap Filling Analysis

In the rainy month period (July-August), one can see the presence of different climate conditions. Hence some days of different cloud condition, having no missing data are selected. Also to verify these days' radiation components at different cloud conditions, time-series plots are shown (see Fig. (5.1)). Cloud condition analysis and quality control guidelines are applied and the quality control results are provided in Fig. (5.2). Here, obviously being a raw radiation database, selected data is found failing in some quality control tests, as identified by the quality control results. After analyzing both plots, it is identified that coherence and tracking error are present at start of each day and during heavy cloud conditions.

Errors are often due to maintenance issues as discussed in chapter 4 (dust on sensors). Hence according to the guidelines, the data correction is suggested in some intervals of GHI and DHI values. Guidelines provided by "C-WET, GIZ" are used for filling these values. Now after this step, the corrected radiation database and quality control plot is provided in Fig. (5.3) and Fig. (5.4). Hence the finalized quality control plot shows, all data passing the quality tests and now database are ready for the next analysis.

Table 5.3: Comparison of Various Gap Filling Approaches

Approach	*Key-points*	*Gaps in radiation gap prediction*
CWET-GIZ	- Can be used for any frequency dataset. - This approach combines relevant gap filling methods. - Here guidelines are controlled by the solar zenith angle. - According to this case-study, approach predicts missed data gap accurately.	- In the double gap filling problem, the proposed approach takes a lot of time; other lighter approaches can be used. - In gap prediction for greater than 3 intervals, sometimes neighboring day values don't follow the same climatic conditions. - In different cloud conditions, approach is not giving consistent results.
Meteorological parameters	- The model predicts good value if supporting meteorological data is measured uniformly. - Approach predicts good hourly averaged values. - Can be used for calculating sub-hourly values, but require same frequency meteorological dataset.	- Parameters have to be updated for every location. -Difficulty in getting reliable data for every location. - Getting sub-hourly meteorological data for any location is difficult. - effects of aerosol, sand, etc. are not included
RMIB	- Used models good for hourly data estimation, provided accurate GHI value is present. - Equation controlled by solar elevation angle and clearness index ratio. - Can be used for estimation of sub-hourly values, but need upgradations in equation parameters. - Best models are provided, for correct estimation of GHI, DNI and sunshine duration.	- For determining constants used in the model, reliable historical accurate dataset is required. - For accurate gap filling, a plot of diffuse ratio with clearness index for different solar elevation angle is required for given location. -Angstrom model and standard model equation, don't give accurate sub-hourly values. -Estimation for clear sky and the fully covered sky is good, but for the mixed cloud, it suffers. - The model fails at large gaps in data.
Curve fitting	- A good approach for filling gaps in all types of cloud condition days. - the polynomial curve fit is better. - For small gaps, at any cloud condition, the approach provides good results. -Can be applied to any type of data set. - Implementation is easy.	- No logic for various cloud conditions -just averages all available values. -High variability in data, predicted value is less accurate. - The model fails at large data gaps.
Satellite Radiation Database	-No Missing Gap and Error Free Database -Hourly average database available for our station	-At gap filling interval, predicted value shows a pattern of data averaging -Maximum frequency data available as of 15-minutes

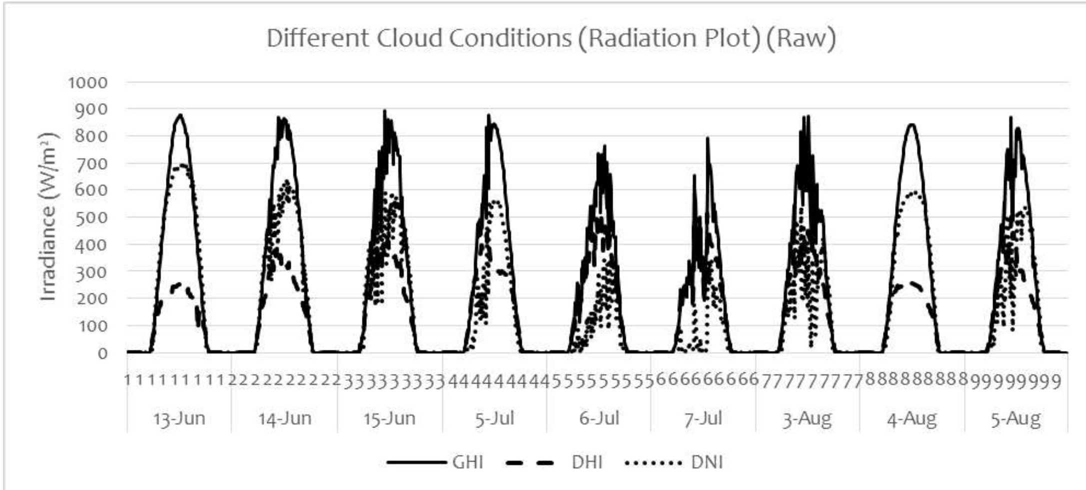


Figure 5.1: Sample Days Selected from Rainy Month, with Different Cloud Conditions

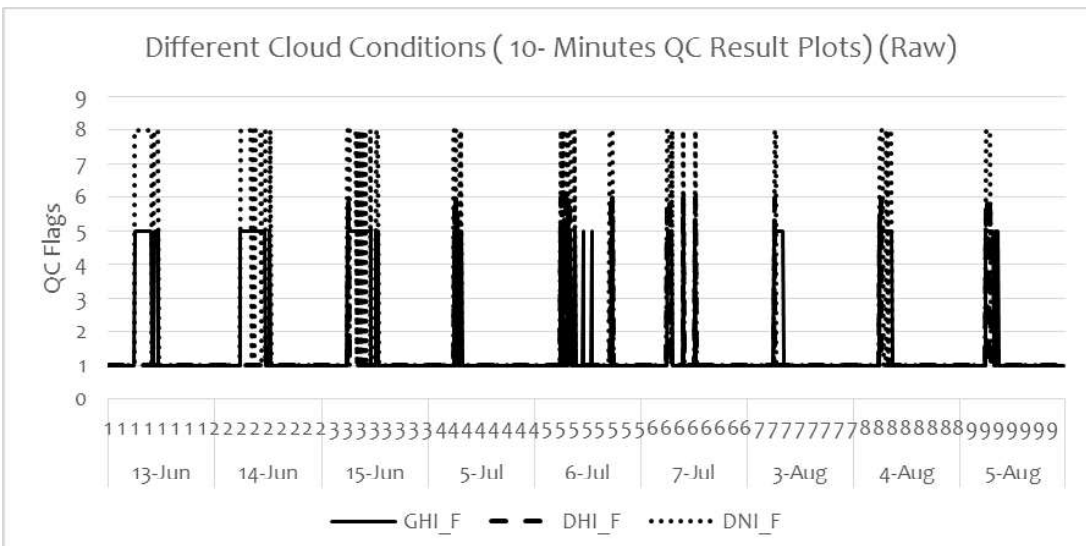


Figure 5.2: Quality Control Result of Above Selected Radiation Database from Rainy Month

5.6.2 Selection of Gap Filling Approach

The objective of this section is to find the relevant gap-filling procedure for different cloud conditions. Artificial gaps in data are created uniformly in the modified database prepared for gap filling analysis. Several types of missing gap conditions are included for each climate condition (at each day's sunshine interval) on selected days. For a better understanding of gap filling results, these days are arranged in a group of different cloud conditions present at the measurement site. At the start of the database, clear sky days are listed, and then some partial cloudy days and at last some heavy clouded days are arranged accordingly (see Fig. 5.5). In this modified database, all missing values are filled by suggested gap filling approaches.

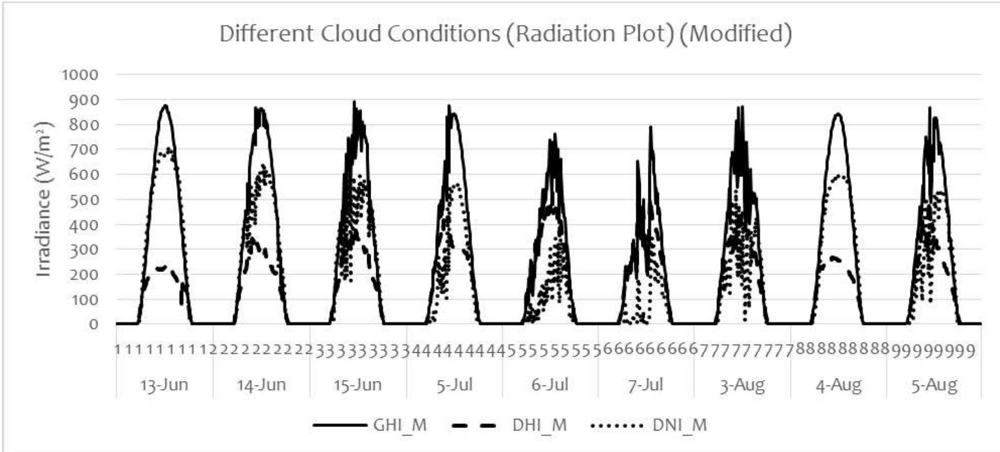


Figure 5.3: Sample Days Selected from Rainy Month, with Different Cloud Conditions (After Correction)

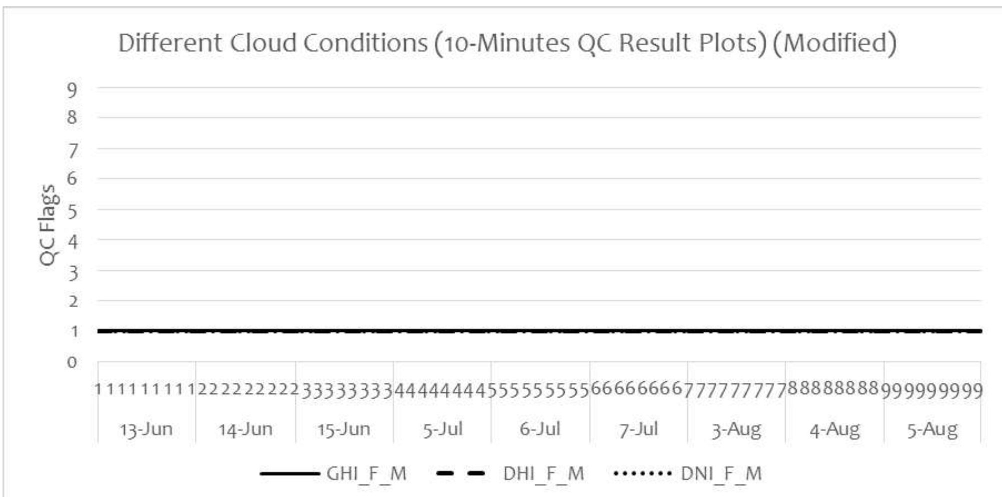


Figure 5.4: Quality Control Result of Above Selected Radiation Database from Rainy Month (After Correction)

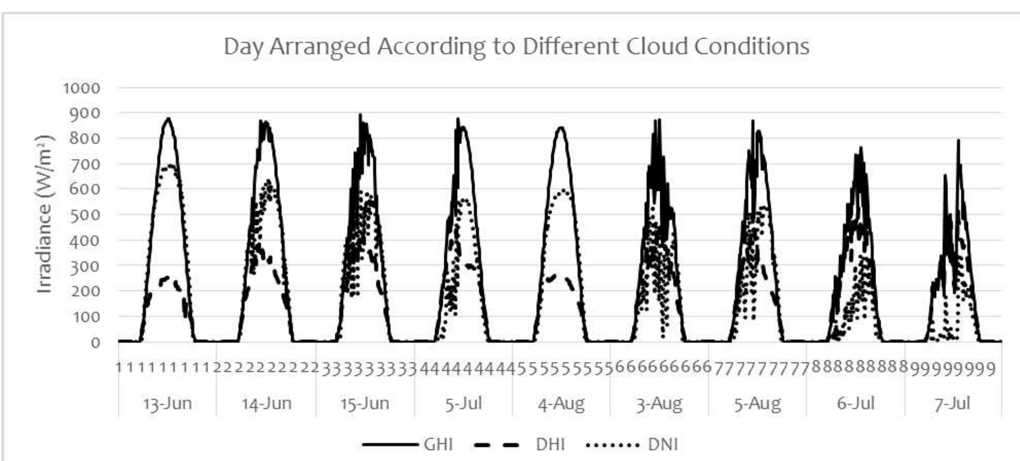


Figure 5.5: Finalized Radiation Database Arranged According to Different Cloud conditions

5.6.3 Gap Filling Approach for Different Cloud Conditions

A modified database having artificially introduced gaps is now used to illustrate the gap filling approaches. The difference between ground-based measurements and predicted values are compared for each approach. Here analysis is done for different cloud conditions and for each radiation component. Based on predicted values, the most economical approach, with minimum input requirements for different gap frequencies and different cloud conditions is suggested.

Cloud condition identification is done using visual analysis of Fig. 5.5. Here it is identified that for the first two days, uniform measured radiation pattern is present between morning and evening values. Some fluctuations in radiation component measurements are seen, but the difference between them is less than 100 (W/m^2). Hence these days fall under clear sky days. This is also validated by plotting their transmittance values. For verifying partial cloud condition days, the limits identified in Table 3.4 are followed. From the modified database, the first three days have GHI variation of 100-200 (W/m^2) and the next two days have a variation of 300 (W/m^2). For heavy cloud condition days, GHI variation of 300-500 (W/m^2) and low GHI and DNI measurement is seen, having less sunshine duration during the days. Here for each radiation component different gap-filling approaches are discussed. Comparisons are done with ground-measured value (removed during the artificial gap creation process). For each case, RMSE values are calculated and these are used for comparing various approaches.

Results for GHI calculation for different cloud conditions are analyzed in Figure (5.6). RMSE value is calculated for each day. For days identified as a clear sky, RMSE obtained by all approaches are 2.31, 6.49, 6.32 and 1.82 ($W/m^2/day$) respectively. Here all approaches provide low RMSE values, but minimum value is for curve fitting and "C-WET, GIZ" approach. Next GHI component analysis at partial cloud condition days is done. RMSE values are in a range of 1.58-12.14, 3.27-12.54, 4.84-15.86 and 1.68-7.01 ($W/m^2/day$) respectively. According to these results, one can conclude that the results obtained are mixed. "Curve fitting" approach seems to be slightly better than others. Due to its simplicity and if the cloud variation is not frequent throughout the day, this approach is suggested for gap filling. Here when comparing gap filling results predicted using curve fitting approach with other applicable approaches, one finds some co-relations which are used to predict the missing values (equations derived using hourly averaged values). Hence the accuracy obtained by these approaches, can't match with actual values fitted by using measured values. For heavily clouded or overcast days, RMSE from different approaches is 4.38, 10.42, 17.39 and 6.62 ($W/m^2/day$). For these days, the variation in global radiation values is high (300-500 (W/m^2)), and the approach suggested by "C-WET, GIZ" provides good results. Hence, due to data missing for long interval and frequent changes in radiation values, best gap filling is only possible by using values of a nearby ground station.

DNI component is the most sensitive radiation component, a slight change in cloud disturbance gives instant radiation fluctuation. Gap filling of direct (DNI) solar radiation component for clear sky days, the RMSE values are given as 2.92, 11.2, 4.66 and 2.83 ($W/m^2/day$) respectively (see in Figure (5.8)). Here minimum RMSE is achieved using "curve fitting" and "GIZ" approach. The approaches use "polynomial" and "physical radiation relationship" equation. Similarly on clear sky day, DNI component also follows a similar path like GHI component and gaps get easily filled by using either near day available values or the derived polynomial equation. Now on a partially clouded day, the RMSE values are in a range of 4.61-19.92, 3.69-15.83, 3.34-8.49 and 1.36-11.01 ($W/m^2/day$) respectively (see Fig. (5.8)). Here the gap prediction results vary at different cloud conditions, and all approaches suffer in predicting the missing value. Therefore, considering all gap filling approaches, curve fitting approach provided good results. It averages the whole day data and suggested equation can fill the gaps automatically. Whereas in heavy clouded condition, theoretically no or less direct

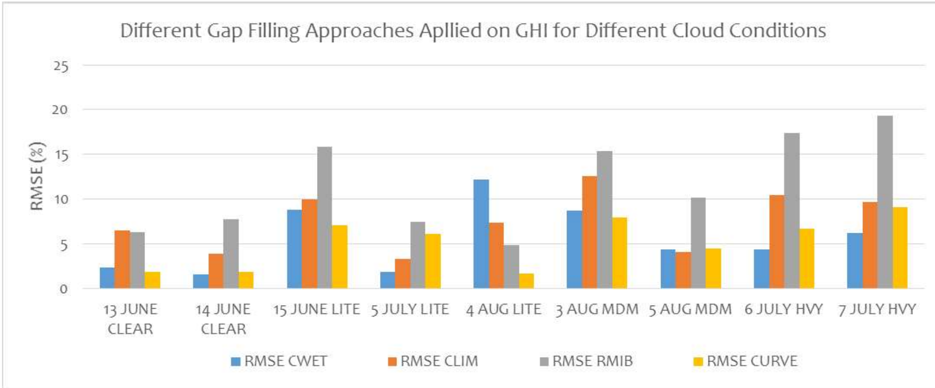


Figure 5.6: Different Gap Filling Approaches Applied on GHI for Different Cloud Conditions

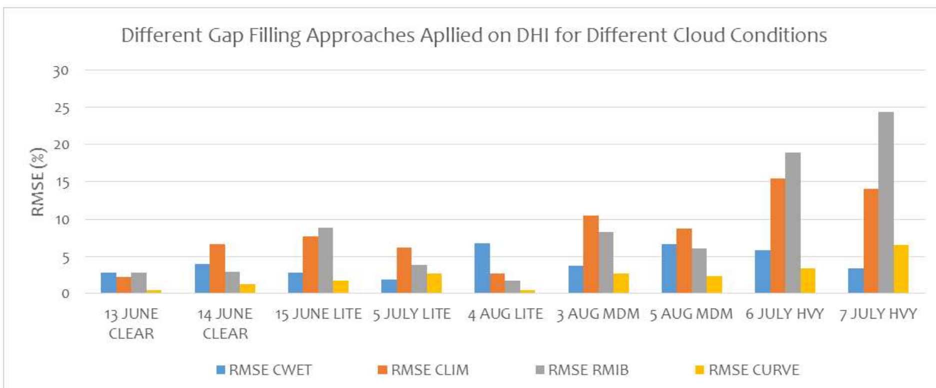


Figure 5.7: Different Gap Filling Approaches Applied on DHI for Different Cloud Conditions

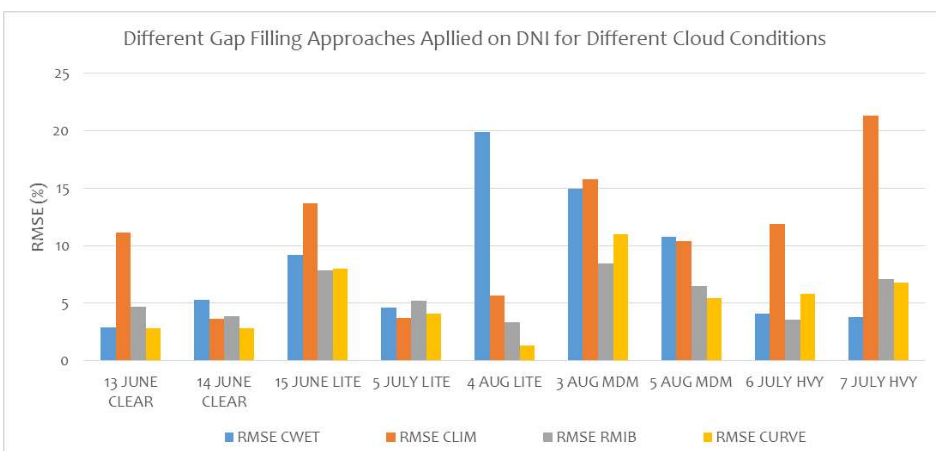


Figure 5.8: Different Gap Filling Approaches Applied on DNI for Different Cloud Conditions

radiation is recorded at the station and presence of heavy clouds makes the step change of DNI values is high. The RMSE for these days are 3.82, 21.35, 7.13 and 6.81 ($W/m^2/day$) respectively and accurate results are achieved, when the physical radiation relationship is used for gap value determination. The radiation value measured by other instruments can only provide us the real picture of current direct radiation value. Here curve fitting doesn't produce good results, due to the presence of mixed cloud conditions at the given location.

Now gap filling of DHI component is the simplest, as one has correctly identified GHI and DNI components values from the above guidelines. However, this component is not having a high step rate as other radiation components, hence a low error is seen while predicting these values. Now on a clear sky day, the RMSE values in DHI estimation are 2.79, 2.27, 2.81 and 0.35 ($W/m^2/day$) respectively as calculated from different gap-filling approaches (see Fig. (5.7)). All models performs well here, hence any approach can be used for gap filling. Now for partially clouded days, the RMSE values are 4.61-19.92, 3.69-15.83, 3.34-8.49 and 1.36-11.01 ($W/m^2/day$) respectively and for heavy clouded conditions day, RMSE values are 3.82, 21.35, 7.13 and 6.81 ($W/m^2/day$) respectively from different approaches. It is also seen that as cloud intensity is increasing, gap prediction ability decreases. This is because the diffuse part of the solar radiation is created by contributions of various factors such as environmental parameters. Theoretically, this component is a subtracted part of GHI and DNI component and if any amount of error is present in these components, DHI value suffers. According to the results, the best approach for filling DHI component is by using the polynomial curve, as the measurement range of this parameter is less (200-400 (W/m^2)) and small variation while prediction is not of significant importance.

5.6.4 Different Techniques of Gap filling Approaches

A list is prepared of all gap filling techniques used for different intervals of missing values in solar radiation database as shown in Table 5.4. Now by using any of these techniques, gap filling can be done. "CLIM" model uses the physical radiation relationship for single gap prediction and for remaining other types of gaps, suggested meteorological relations are used. Similarly, "RMIB" approach uses physical radiation relationship equation for single gap prediction and suggested "Skartveit model" relations for other gaps. "C-WET" approach uses "physical radiation relationship" for single gap prediction, "RMIB" approach for double gap filling, "k_t-k_n" for triple gaps and for greater than that "near station or day" data is used. "Curve fitting", uses linear or non-linear equation for gap filling. Another gap-filling approach is the use of "satellite radiation database", where by comparing of ground and satellite radiation databases missing gaps are filled. Hence, gap values calculated by these approaches are compared with ground-measured values and difference (RMSE) is plotted in Fig 5.9 to Fig 5.12. For comparing each gap-filling approach, (RMSE) values are shown in "bars" and approach giving values closer to ground measured values is shown by "line".

Now in single gap filling, the data missing in radiation database can be in any measured solar radiation component. All approaches used to predict single gap missing values are listed in Table 5.4. One finds that for all types of cloud conditions, good estimates (RMSE) are received when physical radiation relationship and curve fitting approach is used (see Figure 5.9 (a, b and c)). Where for clear sky day, radiation components are available from reliable measuring instruments (no instrument or operational error presence), hence they are directly correlated with each other. However, if in some cases, solar zenith angle is not available for that location and its calculation is also not possible, the value is determined using "k_t-k_n" value. Hence from Figure 5.9 (a, b and c), for single gap filling, guideline (using physical radiation relationship) suggested by "RMIB and C-WET" is found applicable for all cloud condition days.

Table 5.4: Various Gap Filling Approach Applicable for Different Intervals

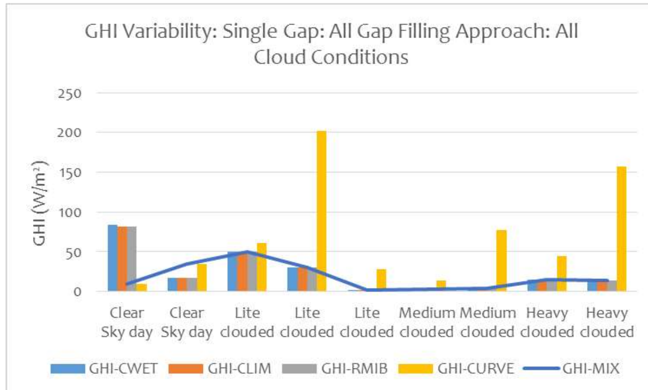
Single gap filling	Double gap filling	Triple gap filling for single period	Triple gap filling for longer period
-Physical radiation relationship -Averaging -k_t-k_n -Curve fitting -Satellite radiation database	-[Skartveit *et al.*,1987] model, then physical radiation relationship -Using meteorological relations. -Curve fitting -Averaging -k_t-k_n -Satellite radiation database	-k_t-k_n (SolarGIS) -Using meteorological relations. -[Angstrom, 1924] and [Skartveit *et al.*,1987] -Curve fitting -Averaging -Near day data -Satellite radiation database	-Averaging -k_t-k_n -Curve fitting -Near day data -Using meteorological relations. -Satellite radiation database

Double gap filling condition is only possible if in time-series for that identified interval, only GHI (reliable) is present and other two radiation components are missing. Here for prediction of DHI or DNI component using present GHI value, "skartveit model" is used with location-specific constants. The results obtained by using the above model (or "diffuse ratio"), for prediction of DHI values, are valid for all types of clouded sky conditions. But for a quick missing value estimation, one can also try other gap filling options like "curve fitting" or "k_t-k_n" (RMIB and C-WET) approach. Now that two radiation components are available, it becomes a problem of single gap filling. Hence "curve fitting" and "physical radiation relationship" equation (RMIB and C-WET), is used here for estimation of the third missing component (see Fig. 5.10 (a, b and c)).

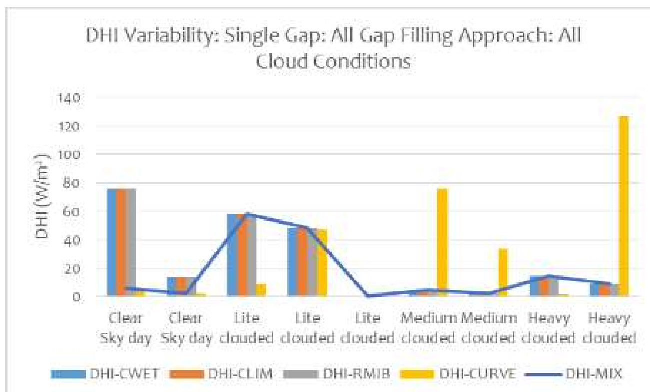
For the triple gap filling problem of short duration (less than three intervals), the approach given by "C-WET" (using "k_t-k_n") is found to predict less RMSE value, as compared with others approaches (see Fig. 5.11 (a, b and c)). By using "C-WET" approach, the present condition of measuring quality is maintained and the best intermediate solution is achieved. These results can also be achieved by using averaging or curve fitting approaches. But if one finds that, measured radiation database is not reliable, then by using a modified "Angstrom" model with "Skartveit" model, all components can be calculated. The results derived using relevant meteorological data is good at a certain level, but not accurate at high-frequency data and for large gaps in databases.

If data are missing for long intervals (few minutes to a single day), for finding these missing values, all feasible approaches are given in Table 5.3. Best option in this case is either use nearby site relevant database or near day with same climate dataset (see Fig. 5.12 (a, b and c)). "C-WET" suggests the use of near day data, as for near days the sun position and climate changes are not so significant, and without any complex calculation, data can be easily filled. Other approaches (provided in Table 5.3) may also work well, but the availability of location specific climate pattern and other relevant meteorological parameters are needed.

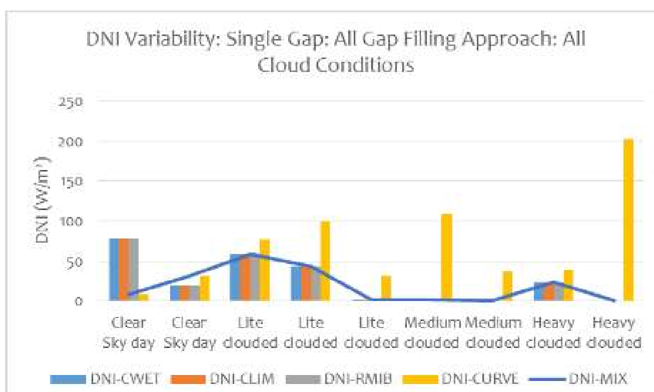
One popular approach used for filling the missing gap is by using satellite radiation database. Missing data of any radiation component (GHI and DNI) can be filled using a satellite database. The available satellite database for any location is of hourly or sub-hourly interval frequency, hence this approach is found helpful for filling data missing during a long



(a)

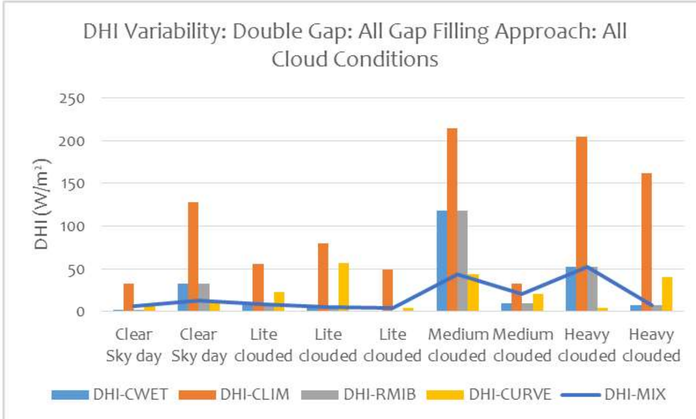


(b)

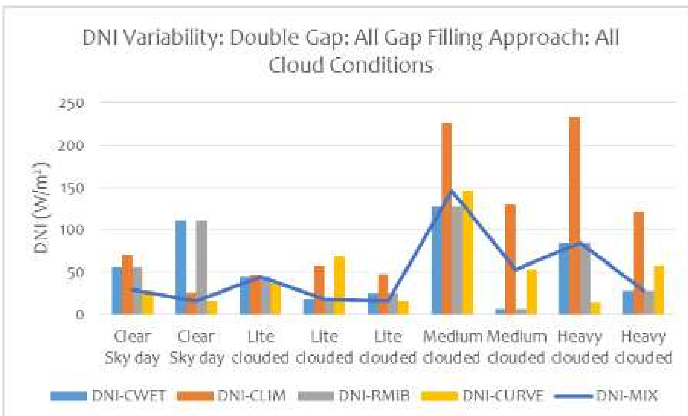


(c)

Figure 5.9: Radiation Components (a. GHI, b. DHI and c. DNI) predicted variability (RMSE) for Single Gap Presence in Database and Different cloud Conditions



(a)



(b)

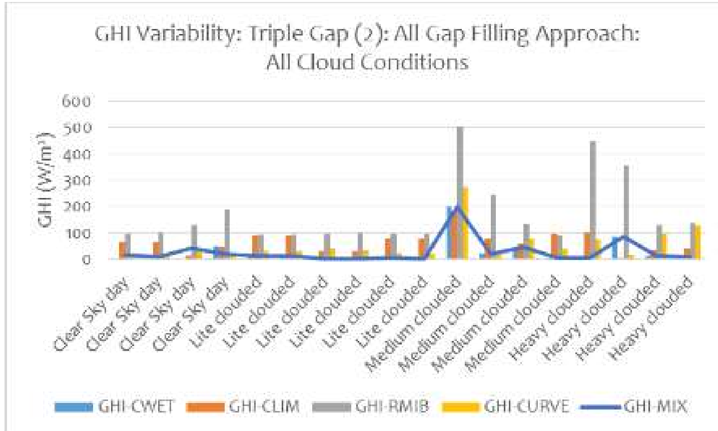
Figure 5.10: Radiation Components (a. DHI and b. DNI) predicted variability (RMSE) for Double Gap Presence in Database and Different cloud Conditions

interval. In addition, if satellite data is compared with nearby site available radiation database, reasonable data prediction can be done. Here one has identified most optimal gap-filling procedure for different types of gaps present, by comparing their RMSE values as in Fig. (5.9) to Fig. (5.12).

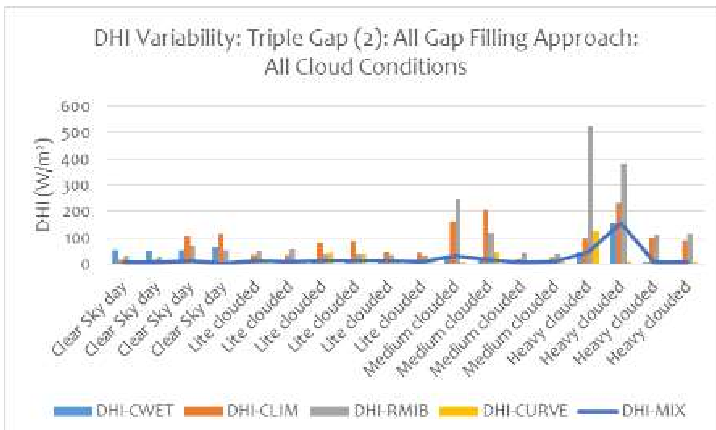
5.6.5 Guidelines for use of Gap Filling Approaches

The main objective of this research work is to find an accurate approach, which is simple to apply and requires less computation for filling the missing gaps in the radiation database. A summary is given in Table 5.5 for the proposed gap filling guidelines for different radiation components at different cloud conditions and frequencies.

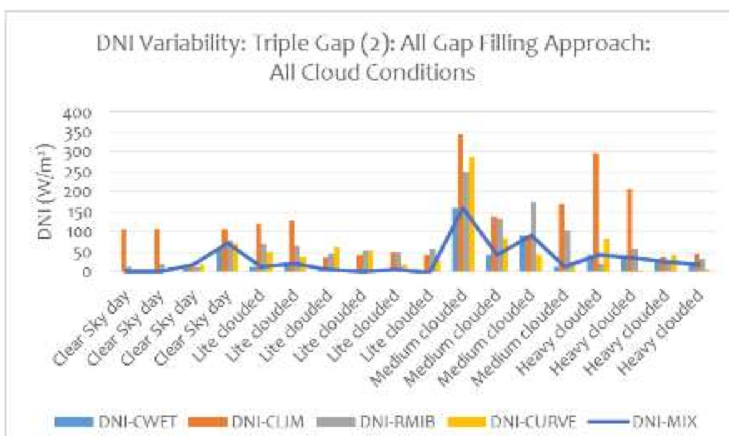
For clear sky day, it is identified that using "physical radiation relationship" equation between radiation components and "curve fitting" approach gives the most accurate gap-filled data. "Curve fitting" approach is also suggested for prediction of (GHI and DNI) radiation components for single gap filling condition. GHI and DNI are the most sensitive radiation components, hence for their prediction "curve fitting" approach is used and after that DHI can be calculated using "physical radiation relationship" equation. For double and triple gap filling (data



(a)

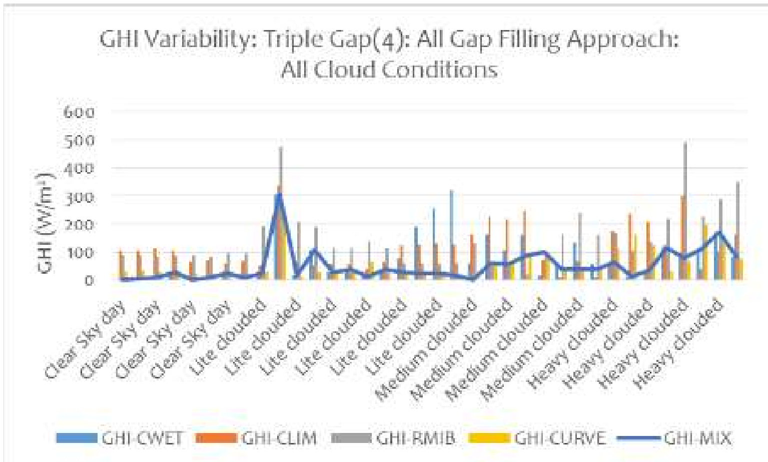


(b)

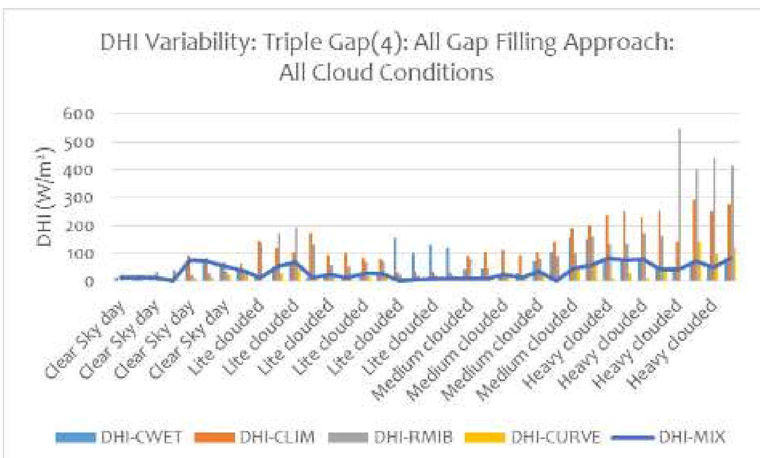


(c)

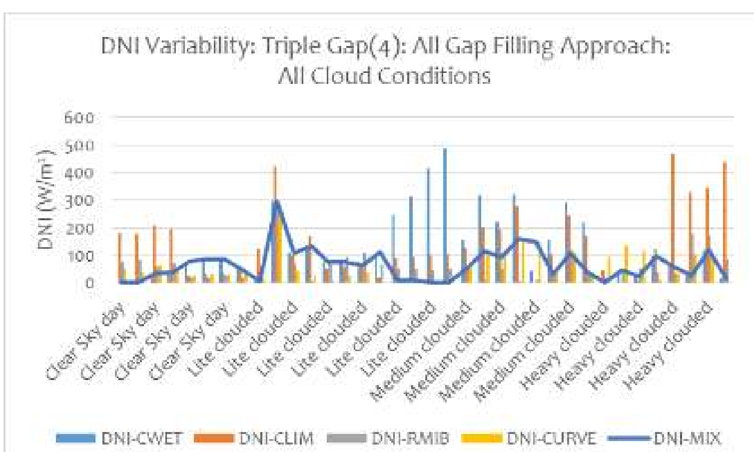
Figure 5.11: Radiation Components (a. GHI, b. DHI and c. DNI) predicted variability (RMSE) for Triple Gap Presence in Database and Different cloud Conditions



(a)



(b)



(c)

Figure 5.12: Radiation Components (a. GHI, b. DHI and c. DNI) predicted variability (RMSE) for High Interval Triple Gap Presence in Database and Different cloud Conditions

missing in one interval) for clear sky days, the gap prediction is accurate by using a "curve fitting" approach, as compared to other gap filling approaches. But if missing intervals are high (few minutes to hours), then all approaches fail due to unavailability of a standard best fit model for the given location. Hence, the suggested approach is to use nearby radiation day database, having the same cloud conditions for that period. Where the identification of correct cloud condition day is done by using guideline given in Table 3.4.

Table 5.5: Proposed Gap Filling Guidelines for Filling any Interval Data

		GHI	*DHI*	*DNI*
Clear sky	Single gap	Curve fitting	Curve fitting	Curve fitting
	Double gap		Curve fitting	Curve fitting
	Triple gap (<3)	Curve fitting	Curve fitting	Curve fitting
	Triple gap (>3)	Previous day	Previous day	Previous day
Light cloud	Single gap	Physical R-ship	Physical R-ship	Physical R-ship
	Double gap		Physical R-ship	K_n ratio
	Triple gap (<3)	K_t ratio	Physical R-ship	K_n ratio
	Triple gap (>3)	Previous day	Previous day	Previous day
Medium cloud	Single gap	Physical R-ship	Physical R-ship	Physical R-ship
	Double gap		Curve fitting	Curve fitting
	Triple gap (<3)	Curve fitting	Curve fitting	Curve fitting
	Triple gap (>3)	Previous day	Previous day	Previous day
Heavy cloud	Single gap	Physical R-ship	Physical R-ship	Physical R-ship
	Double gap		Skaweith model	Physical R-ship
	Triple gap (<3)	K_t ratio	Physical R-ship	K_n ratio
	Triple gap (>3)	Previous day	Previous day	Previous day

For identification of light clouded sky conditions, one can spot small variations in the measured radiation components (when compared with clear-sky day). The possible reasons behind this variation can be many, but for the well-maintained station, it is due to the presence of thin and small clouds present in the sky. Days having step variations of 50-100 (W/m^2) in GHI values generally fall under this group. Due to these fluctuations, the "curve fitting" approach predicts a less accurate value for that time interval. Hence direct "physical radiation relationship" is the best method for filling these gaps for single and double gap filling locations. But if all components are missing (triple gap filling case) then "k_t-k_n" gives best-predicted value for GHI and DNI values and after that "physical radiation relationship" equation is applied for determining the missing value. For data missing for greater than three intervals, find a similar cloud condition for a nearby day and missing gaps are filled using that information.

In medium clouded sky conditions, step variation in GHI component is about 300 (W/m^2). The presence of frequent heavy cloud, effects the solar radiation falling at the measurement location. Under this condition for a single gap problem, using physical radiation relationship equation gives the best solution. For double and triple gap filling cases, due to unavailability of radiation values and frequent cloud movements, the components of the solar radiation are not related easily. If for a whole day, the cloud conditions are the same, the "curve fitting" approach averages the whole day values and results are calculated. Here, if the condition is only present for a small portion of time, then "k_t-k_n" is used for all radiation parameter estimation. If data is missing for greater than three intervals, find a near day with similar cloud condition.

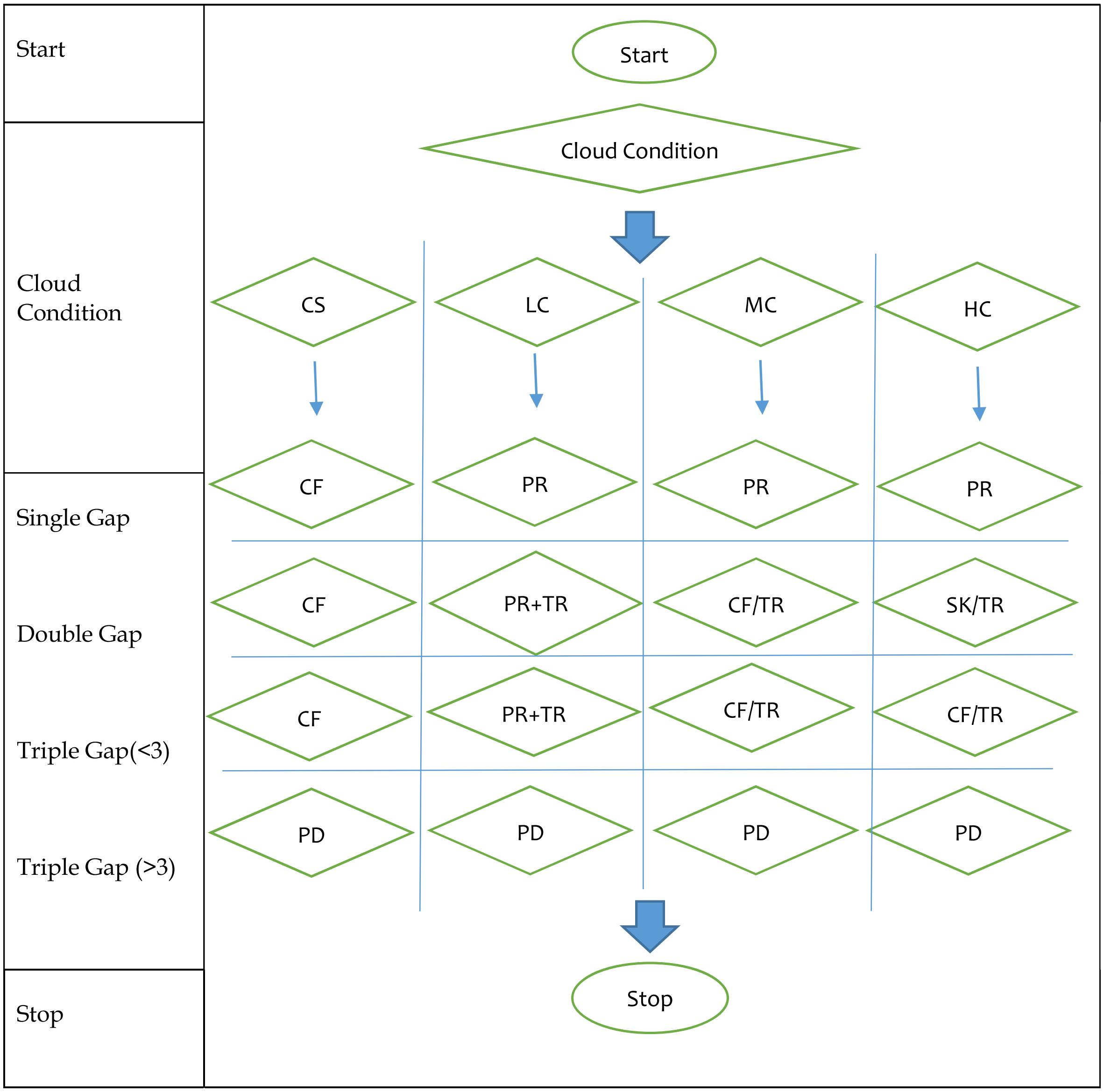


Figure 5.13: Flowchart showing Proposed Gap-Filling Procedure for Different Cloud Conditions

The flowchart showing the proposed gap-filling procedure for different cloud conditions is given in Fig. 5.13. Where "CS, LC, MC and HC" (Clear sky, Light Clouded, Medium Clouded and Heavily Clouded) shows different cloud conditions. These are identified during quality control analysis and accordingly applied to the radiation database. For filling the missing gaps, methods used are: Curve fitting (CF), Physical Radiation Relationship (PR), Transmittance Ratio (TR), Skartveit model (SK) and Previous Day (PD). Hence by using this flowchart, gap filling will be done in the next chapter.

5.7 SUMMARY

In this chapter, according to the missing interval, gaps are identified as single, double, (for less than 3 intervals) and triple (for greater than three interval) for each radiation component. For filling these missing values, guidelines are suggested.

Approaches which are found applicable for high-frequency radiation dataset are discussed. These are illustrated with a database selected which is free from missing gaps, and different frequency artificial gaps are uniformly introduced in it. After that, the suggested gap filling approaches are applied to this modified database and missing values are calculated and compared with other approaches. To achieve quality of ground-based measurements analysis is done based on different cloud conditions, gap identifying approaches and gap frequency presence. Also, each calculated value is compared with ground-measured values and for all comparisons "RMSE" is calculated. For a clear sky day (except for triple gaps having greater than 3 intervals), curve fitting approach is found to the best. For lightly clouded sky due to some cloud disturbance, GHI and DNI are calculated by "k_t-k_n" ratio and DHI by "physical radiation relationship equation". Now in medium clouded condition, in the presence of dense clouds use of "k_t-k_n" ratio and for cloud moving condition, use of "curve fitting" for GHI and DNI calculation is suggested. Now, remaining DHI value is calculated by using "physical radiation relationship equation". Finally, for heavy cloud conditions, physical radiation relationship equation estimates good radiation component values at single gaps. In double and triple gaps, "k_t-k_n" ratio and "skartveit" model is used for all radiation component values. Similarly, for triple gap filling (data missing for greater than 3 interval), only selection of near day (with similar cloud condition day) can fill these missing values.

For further accurate predictions, some computational models are available, but these are not included in this work, as this makes the data analysis system complex and some historical data is also required for further work. Using all these guidelines, the available database will be carefully analyzed in the next chapter, and its relevance for solar resource assessment is examined.

….