Real-time regional carbon emissions intensity model detailed description

TECHNICAL REFERENCE DOCUMENT
Australia's National Science Agency
Energy Systems Transition
Energy

Version History

Version	Date	Description
1.0	20 Aug 2026	First release

Citation

Goldsworthy M, Aryai V, McCulloch J (2026) Real-time regional carbon emissions intensity model Technical Reference. CSIRO, Australia.

Copyright

Important disclaimer

Acknowledgement of Country

CSIRO acknowledges the Traditional Owners of the lands, seas and waters of the area that we live and work on across Australia and pays its respects to Elders past and present. CSIRO recognises that Aboriginal and Torres Strait Islander peoples have made, and will continue to make, extraordinary contributions to Australian life including in cultural, economic, and scientific domains.

Contents

1 Introduction

The real-time carbon emissions intensity product for the NEM provides seven data streams - the regional emissions factors for the 5 NEM regions (Qld, NSW, Vic, SA & Tas) and a global quality flag as scalar timeseries, as well as a more detailed timestamped metrics data stream. This document provides a technical description of the method used to calculate regional-level emissions intensity factors for the National Electricity Market at 5-minute intervals. The method is based on a foundation paper by Aryai and Goldsworthy (Aryai & Goldsworthy, 2022), with several improvements including the addition of interconnector flow losses, use of variable generator emissions factors and the inclusion of behind the meter solar generation.

Real time execution of the algorithm below occurs every 5 minutes. With each execution, the last 3 hours of time-series data is fetched, calculations are performed, and results written. Written results will overwrite earlier results. Once per day, the algorithm fetches the last day of time-series data as well as the latest generator emissions factor spreadsheet and again redoes the calculations and overwrites emissions intensity values. This overlapping re-write results strategy is employed because AEMO employ a QA/QC process that does result in updates of the source data after it is initially published. This approach allows results to be available as early as possible, but users should be aware that recent results can change after initial publication. From time-to-time historical results may also be updated however users will be notified prior to any changes being made.

Internal calculations are based on NEM time (GMT+10) but timeseries outputs are reported in UTC time.

2 Detailed algorithm description

The first step of the algorithm is to fetch required data. Timeseries data list in Table 1 is ingested into CSIRO's Data Clearing House platform (DCH) from AEMO. Distribution Unit Identifier's (DUID's) in the power flow (SCADA) data are then matched with DUID's in the EF data and any SCADA data without matching EF values is removed. This may mean that some actual generation is removed from the calculation (this removal and uncertainty is addressed below).

SCADA data includes data from generators (both scheduled, semi-scheduled and unscheduled), scheduled loads and bi-directional units (BDU's). The 'type of unit' is used to ensure sign conventions are correct, where input generator and scheduled load data values are positive, while data from bi-directional units is positive for generation and negative for load.

Two important concepts in calculating emissions intensity factors are "Total demand" and "Demand and nonschedgen". Both are specific quantities defined by AEMO. Total demand (TD) represents "the underlying forecast demand at the Regional Reference Node that is met by local generation from scheduled and semi-scheduled generating units and scheduled BDUs (Bi-directional Units[1]) plus interconnector imports, excluding the local demand of scheduled loads and scheduled BDUs and the allocated interconnector losses, but including the demand met by Wholesale Demand Response (WDR)"[2]. "Demand and nonschedgen" (DNSG) is closely related but also includes non-scheduled generation, scheduled loads and allocated interconnector losses while excluding WDR. Emissions values are calculated using a methodology based on DNSG.

Using DNSG instead of TD means that interconnector flow losses are automatically included which simplifies the equations. Bidirectional unit charging is removed from the SCADA data by setting negative SCADA values to zero. Values from schedule loads are automatically excluded as they don't have a matching EF value in the emissions factor dataset. Including the demand met by non-scheduled generation and the demand caused by bi-directional unit charging removes the error introduced by any mismatch between SCADA based generation and demand values. However, inclusion of scheduled loads in the demand side (absent in the SCADA data due to the emissions factor matching) may introduce a small error. This error is expected to be much smaller than that caused by neglecting non-scheduled generation.

DNSG also excludes small-scale solar generation. That is, behind-the meter generation effectively lowers the demand visible to the network. Hence, the component of small-scale generation in region *r* that is exported (sent out) to the grid $(1 - \sigma_r) \times SG_r$ is added to DNSG to compute the underlying residual demand D_r in the region.

$$D_r = DNSG_r + (1 - \sigma_r) \times SG_r \qquad (1)$$

Here σ_r is the average self-consumption fraction of all small-scale solar generation in region *r*. In general, σ_r is an (unknown) variable that changes with time. However, it is reasonable to approximate it as a function of hour of the day and season for a given region[3].

Estimates of SG_r are obtained from AEMO via the Public-Rooftop-PV-actual field. 30-minute values are linearly interpolated to the 5min interval. The exported component of small-scale PV generation is also added to the generator SCADA data as a 'dummy' generator per region. Prior to Aug 2016 rooftop PV data is not available and is assumed to be zero.

Because of potential excluded generation (due to no matching EF or other data discrepancies), the sum of all generation does not always precisely match the sum of all demand values overall all regions. The difference between the total NEM-wide demand and the generation at a given timestep is calculated as the 'scada_data_missing_pct' and the numbers of unmatched DUID's and the percentage of negative SCADA values are also recorded.

Although the timeseries data is nominally aligned at every 5 min interval, data gaps occasionally occur, and so any time intervals with either no remaining SCADA values or no regional demand and interconnector flow values are then removed. Emissions intensity values are calculated for the remaining timesteps. Hence, there may be missing values in the computed emissions intensity timeseries.

[1] A bi-directional unit is a unit that can be both a generator and load, the most common example is a battery energy storage system.

[2] AEMO, 2024 Demand Terms in the EMMS Data Model v16.0, Australian Energy Market Operator.

[3] The values currently used for σ_r vary by time of day and season based on internal CSIRO datasets. This model is planned to be updated in a future release.

To account for missing generator data, for each timestep remaining SCADA data is uniforming scaled by a scaling factor *SF* calculated from the sum of Demand across all regions for the timestep divided by the sum of generator SCADA data G_{base} for the timestep;

$$SF = \frac{\sum D_r}{\sum G_{i,base}} \qquad (2)$$

$$G_i = SF \times G_{i,base} \qquad (3)$$

The demand fraction *DF* is a matrix of size $N_r \times N_r N_f$ representing the fraction of demand in region *r1* that is met by generation with fuel type *f* in region *r2*. That is, each row corresponds to a demand region, and each column corresponds to generation region-fuel combination.

DF is obtained by solving the linear system in Eq. 4.

$$\mathbf{A^T} \times \mathbf{DF} = \mathbf{B} \qquad (4)$$

$$\mathbf{A} = \begin{bmatrix} D_1 + \sum_n F^+_{1\to n} & -F^+_{1\to 2} & \dots \\ F^-_{1\to 2} & D_2 + \sum_n F^+_{2\to n} & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} G_{1,1} & G_{1,2} & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & G_{2,1} & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & G_{r,f} \end{bmatrix}$$

Where:

D_r: Demand in region *r* (including behind the meter generation & allocated losses)

$F^+_{1\to n}$: Positive component of flow from region 1 to region *n*. i.e. $F^+_{1\to n} = \max(0, F_{1\to n})$

$F^-_{1\to n}$: Negative component of flow from region 1 to region *n*. i.e. $F^-_{1\to n} = \min(0, F_{1\to n})$

$L_{1\to n}$: flow loss on interconnector between region 1 and region *n*

$F_{L,1\to n}$: inter-regional loss coefficient for region 1 for interconnector flow 1→*n*

$G_{r,f}$: generation in region r by fuel type f

Note negative flow loss values do occur in the data and arise due to the approximate method used by AEMO to estimate the static inter-regional loss factors.

In theory, the sum of each row of DF should equal 1 since the sum of total demand and losses in each region must be met by some combination of generation-fuel types across all regions. To ensure this condition is met, DR is rescaled by the sum of the columns. That is,

$$DF_{r,f} = \frac{DF_{r,f}}{\sum_{rf} DF_{r,rf}} \qquad (5)$$

The final regional emissions intensity value is calculated from

$$\mathbf{EI} = \varepsilon_{\mathbf{c}} \mathbf{DF} \times \mathbf{EF_w} \qquad (6)$$

$\mathbf{EF_w}$ is the $(N_r N_f \times 1)$ weighted emissions factor vector calculated with each element equal to the generation weighted sum of all the generator emissions factors in a given region and fuel type. That is, each element of EF_w is calculated as:

$$EF_{r,f} = \frac{\sum G_i \times EF_i}{\sum G_i} \qquad (7)$$

Where EF_i is the emissions intensity factor of generator *i* at the time interval being considered (generator emissions intensity factors do change over time) and G_i is the SCADA generation value for generator *i*. The sum in this equation is over all the generators with a given region and fuel type.

The correction factor $\varepsilon_{\mathbf{c}}$ is applied to account for the emissions associated with the generation required to supply the interconnector flow losses in the region and is given by:

$$\varepsilon_{\mathbf{c}} = \frac{DNSG_r}{DNSG_r - \sum_n \left[F_{L,r\to n} \left|L_{r\to n}\right| + \left(1 - F_{L,r\to n}\right) \left|L_{r\to n}\right|\right]} \qquad (8)$$

Flow loss allocation data between region pairs is available from mid-2024 for regulated (non MNSP) interconnectors. Prior to that date, and for MNSP interconnectors, only net interconnector flow loss is available, and loss allocation is done based on annual-fixed loss factors for each interconnector. Note the effect of using fixed loss factors is small as the correction factor ε_c is typically <1%.

Data quality for each timestep is calculated based on SF and ranges from 0 (no result) to 1 (highest quality). Quality is zero if the condition number of the linear solver *r* > 10^{-9}.

$$Q = \begin{cases} \exp\left(-2|1 - SF|\right), & \text{if } r > 10^{-9} \\ 0, & \text{otherwise} \end{cases} \qquad (9)$$

A worked example of applying Eqs. 4→8 to a simplified 3 region - 2 fuel system is given at the end of the document.

3 Data sources

3.1.1 Time-series data

Time-series data listed in Table 1 is sourced from public AEMO data available via NEMWEB. This data is ingested in real-time to the DCH platform. From there, the emissions calculation algorithm fetches the data in chunks based on the analysis period.

Table 1. List of time-series data streams

Data	Symbol	Short description	Detailed description	Source
DEMAND_AND_NONSCHEDGEN	*DNSG*	Regional demand	Underlying demand met by local generation from scheduled, semi-scheduled and unscheduled generating units, scheduled BDUs and scheduled loads, plus interconnector imports, and allocated loses, but excluding demand met by WDR. Logged in 5-minute intervals. Units MW.	(AEMO, DispatchIS Reports, 2026)
METEREDMWFLOW	*F*	Metered interconnector flow	Power-flow through NEM interconnectors (N-Q-MNSP1, NSW1-QLD1, T-V-MNSP1, V-S-MNSP1, V-SA, and VIC1-NSW1) metered flow from SCADA, logged in 5-minute intervals. Units MW.	(AEMO, DispatchIS Reports, 2026)
MWLOSSES	L	Calculated overall interconnector flow losses	MW power loss through NEM interconnectors (N-Q-MNSP1, NSW1-QLD1, T-V-MNSP1, V-S-MNSP1, V-SA, and VIC1-NSW1) logged in 5-minute intervals. Units MW.	(AEMO, DispatchIS Reports, 2026)

Data	Symbol	Short description	Detailed description	Source
FROM_REGION_ MW_LOSSES TO_REGION_MW_ LOSSES		Calculated allocated interconnector flow loss	MW power loss allocation to regions for non-MNSP interconnectors (NSW1-QLD1, V-SA, and VIC1-NSW1) logged in 5-minute intervals. Units MW.	(AEMO, DispatchIS Reports, 2026)
ROOFTOP_PV_ ACTUAL	*SG*	Estimated generation at the interval end	Estimated behind the meter solar generation at 30 min intervals. Units MW.	(AEMO, Rooftop PV actual, 2026)
SCADAVALUE	*G*	Instantaneous MW reading from SCADA at the start of the Dispatch interval	Real-time SCADA data for scheduled, semi-scheduled and non-scheduled generators and bi-directional units recorded in 5-minute intervals. Units MW.	(AEMO, Dispatch SCADA, 2026)
Generator emissions factors – real-time code	*EF*	Generator specific emissions factor	Generator specific emissions factor, updated daily. Units $kgCO_2/kWh$.	(AEMO, CDEII emissions factors, 2026)
Generator emissions factors – historical analysis	*EF*	Generator specific emissions factor	Generator specific emissions factor, updated daily. Units $kgCO_2/kWh$.	(AEMO, MMSDM Archive, 2026)

3.1.2 Non-timeseries data

Table 2. List of non-time-series data used

Data	Symbol	Description	Source
Inter-regional loss factors	F_L	Loss factor for nominal 'from' region. Units: none.	(AEMO, Marginal Loss Factors (various years), 2026)
Behind the meter PV self consumption fraction	σ_r	Regional PV average self-consumption estimates by month and hour of the day. Units: none.	(Goldsworthy, 2025)
Renewable generator emissions factors	*EF*	Wind: 0.011 $kgCO_2/kWh$ Solar: 0.045 $kgCO_2/kWh$ Hydro: 0.024 $kgCO_2/kWh$ Biomass: 0.23 $kgCO_2/kWh$	(Schlömer S., 2014)

4 Output quantities

Primary timeseries outputs are listed in Table 3 and consist of the regional emissions intensity estimates and an overall data quality flag at 5-minute intervals. Secondary outputs that may be used as diagnostics are listed in Table 4. These outputs are written as document stream timestamped to the algorithm run time.

Table 3. Primary outputs at 5 minute intervals

Symbol	Description	Units	Data stream
EI	Regional emissions intensity	gCO_2/kWh	`regional_global_emissions.REGION` where REGION = nsw, qld etc.
Q	Overall data quality flag	-	`regional_global_emissions.quality.nem`

Table 4. Diagnostic outputs written to metrics.nem document stream each time the algorithm is called

Parameter	Units	Description
runtime	string	Time of algorithm execution (GMT+10)
starttime	string	First timestamp updated (GMT+10)
endtime	string	Last timestamp updated (GMT+10)
scada_duid_count	Count	The number of unique DUID found in the scada data

Parameter	Units	Description
timestep_count	Count	Number of timesteps in total demand & interconnector data
belowzero_pct	Percent	Percentage of scada values that are negative
unmatched_duid_count	Count	The number of DUIDs in the scada data for which no matching DUID was found in the emissions factor data.
unmatched_gen_pct	Percent	The percentage of total generation from the unmatched DUIDs across all regions.
unmatched_gen_metadata_count	Count	The number of DUIDs in the emissions factor data for which there was no matching DUID in the scada data.
fueltype_count	Count	The number of unique fuel types found in the scada data.
region_count	Count	The number of regions in the scada data that were matched with regions specified in the configuration file.
scada_demand_meandiff_pct	Percent	Mean percentage relative difference between the sum of scada generation and total demand across all regions. Positive values correspond to total demand < total generation.
scada_demand_maxdiff_pct	Percent	Maximum percentage difference between the sum of scada generation and total demand across all regions.
scada_demand_missingtimestep_pct	Percent	The percentage of timesteps where there is no scada data but there is total demand & interconnector data.
failed_pct	Percent	The percentage of timesteps where the solver was ill-conditioned OR where there was no scada data at all (i.e. where quality=0)
mean_quality	-	Average data quality
computation_time	sec	Overall calculation time

5 Simple worked example of calculation method

```
flowchart LR
    R1[Region 1] -->|F1→2| R2[Region 2]
    R3[Region 3] -->|F3→1| R1
    R3 -->|F3→2| R2
```

Figure 1. Simple 3-region worked example network architecture.

Table 5. Worked example region demand and generation values

Region	1	2	3
Total demand (MW)	5	10	12
Coal generation (MW)	3	5	12

Region	1	2	3
Coal generator emissions factor ($kgCO_2/kWh$)	0.9	0.95	0.85
Solar generation (MW)	1	1	8
Solar generator emissions factor ($kgCO_2/kWh$)	0.1	0.05	0.05

Table 6. Worked example interconnector values

Connection	1-2	1-3	2-3
Flow across interconnector (MW)	3	-5	-2
Flow loss (MW)	1	1	1
Interregional loss factor (-)	0.5	0.5	0.5

$$\mathbf{A} = \begin{bmatrix} (5+0.5\times 2)+3 & -3 & 0 \\ 0 & (10+0.5\times 2) & 0 \\ -5 & -2 & (12+0.5\times 2)+5+2 \end{bmatrix} = \begin{bmatrix} 9 & -3 & 0 \\ 0 & 11 & 0 \\ -5 & -2 & 20 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 12 & 8 \end{bmatrix}$$

$$\mathbf{DF} = \left(\mathbf{A}^T\right)^{-1}\mathbf{B}$$

$$= \begin{bmatrix} \frac{1}{3} & \frac{1}{9} & 0 & 0 & \frac{1}{3} & \frac{2}{9} \\ \frac{1}{11} & \frac{1}{33} & \frac{5}{11} & \frac{1}{11} & \frac{1}{5} & \frac{2}{15} \\ 0 & 0 & 0 & 0 & \frac{3}{5} & \frac{2}{5} \end{bmatrix}$$

Thus:

$$\mathbf{EI} = \varepsilon_{\mathbf{c}} \times (\mathbf{DF} \times \mathbf{EF_w}) = \varepsilon_{\mathbf{c}} \times \begin{bmatrix} 0.606 & 0.698 & 0.53 \end{bmatrix}$$

$$\varepsilon_{\mathbf{c}} = \begin{bmatrix} \frac{6}{6-1} & \frac{11}{11-1} & \frac{12}{12-1} \end{bmatrix} = \begin{bmatrix} 1.2 & 1.1 & 1.083 \end{bmatrix}$$

$$\mathbf{EI} = \begin{bmatrix} 0.727 & 0.768 & 0.574 \end{bmatrix}$$

Note that the correction factor applied in this example is much larger than would typically be the case.

Works Cited

AEMO. (2026). *CDEII emissions factors.* Retrieved from https://nemweb.com.au/Reports/Current/CDEII/CO2EII_AVAILABLE_GENERATORS.CSV

AEMO. (2026). *Dispatch SCADA.* Retrieved from https://nemweb.com.au/Reports/Current/Dispatch_SCADA/

AEMO. (2026). *DispatchIS Reports.* Retrieved from https://www.nemweb.com.au/Reports/Current/DispatchIS_Reports/

AEMO. (2026). *Marginal Loss Factors: Financial Year 2026-27.*

AEMO. (2026). *MMSDM Archive.* Retrieved from https://nemweb.com.au/Data_Archive/Wholesale_Electricity/MMSDM/

AEMO. (2026). *Rooftop PV actual.* Retrieved from https://nemweb.com.au/Reports/ARCHIVE/ROOFTOP_PV/ACTUAL/

Aryai, V., & Goldsworthy, M. (2022). Controlling electricity storage to balance electricity costs and greenhouse gas emissions in buildings. *Energy Informatics, 5(1)*, 1.

Goldsworthy, M. (2025). Solar PV self-consumption trends in Australia. *Solar Energy Advances, (5)*, 7.

Schlömer S., e. a. (2014). *Climate change 2014: Technology-specific Cost and performance parameters.* IPCC.