



Metodologia

L'estimació de les emissions de CO₂ s'ha realitzat mitjançant models de regressió lineal múltiple construïts a partir de dades històriques i variables explicatives associades a l'activitat dels sectors analitzats. La metodologia combina selecció automàtica de variables, validació estadística dels models i projecció sota diferents escenaris futurs.

El càlcul s'ha portat a terme mitjançant un programari en R específic disponible sota demanda.

Preparació i tractament de dades

En primer lloc, s'han carregat les dades històriques (2012-2023) corresponents a:

- les variables explicatives o predictors (consum de gasolina o gasoil, gasoli B, fueloil, gas, emissions sector directiva, etc.),
- i les variables objectiu, corresponents a les emissions de CO₂ registrades als principals sectors i total a l'inventari oficial

Adicionalment, s'han incorporat diversos escenaris futurs a partir de dades oficials d'accés obert corresponents als anys 2024 i 2025 (CORES e ICAEN). Cada escenari conté projeccions futures dels predictors i permet generar estimacions diferenciades de les emissions.

Construcció dels models lineals

Per a cada variable d'emissions s'ha ajustat un model de regressió lineal múltiple de la forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

on:

- y representa les emissions de CO₂,
- x_i són les variables explicatives,
- i β_i els coeficients estimats pel model.

Inicialment, cada model incorpora tots els predictors disponibles.



Selecció automàtica de variables

Per evitar sobreajustaments i reduir la complexitat dels models, s'ha aplicat un procediment automàtic de selecció de variables mitjançant l'algorisme *stepwise* bidireccional.

La selecció es basa en el criteri BIC (*Bayesian Information Criterion*), que penalitza els models excessivament complexos i afavoreix solucions més equilibrades. Aquest procediment elimina els predictors que no aporten capacitat explicativa significativa i conserva únicament les variables més rellevants per a cada model.

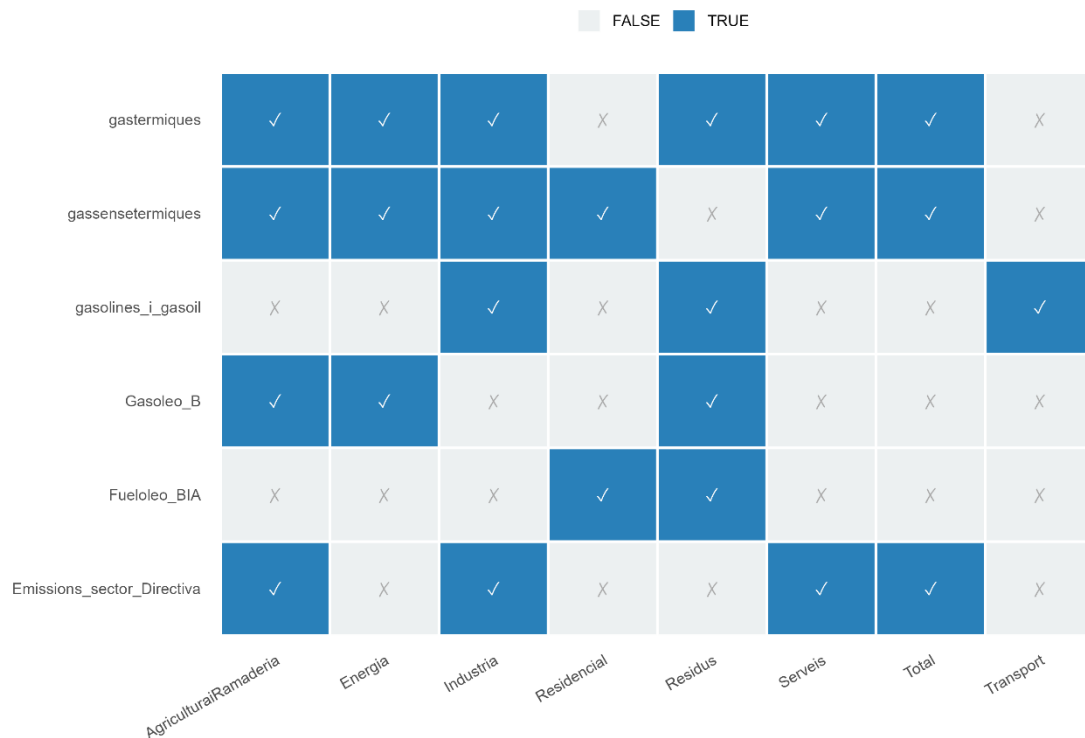


Figura 1 Predictors retinguts per variable (selecció per AIC).

Avaluació estadística dels models

La qualitat dels models ajustats s'ha avaluat mitjançant diferents indicadors estadístics:

- coeficient de determinació (R^2),
- coeficient de determinació ajustat,
- RMSE (*Root Mean Squared Error*),
- AIC (*Akaike Information Criterion*),
- BIC (*Bayesian Information Criterion*).

El coeficient R^2 s'ha utilitzat com a indicador principal de la capacitat explicativa del model:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

on SS_{res} correspon a la suma de quadrats dels residus i SS_{tot} a la variabilitat total observada.

El paràmetres corresponents als models de cada sector i total són:

| AIC | AICc | BIC | R2 | R2_adjusted | RMSE | Sigma | Variable |
|----------|----------|----------|----------|-------------|----------|----------|-----------------------|
| 151,7523 | 161,7523 | 154,1768 | 0,971591 | 0,960938179 | 88,89066 | 108,8684 | Energia |
| 172,3705 | 189,1705 | 175,28 | 0,963125 | 0,942053152 | 193,092 | 252,8168 | Industria |
| 143,4513 | 149,1656 | 145,391 | 0,826401 | 0,787822936 | 68,36545 | 78,94162 | Residencial |
| 149,3047 | 166,1047 | 152,2141 | 0,596789 | 0,366382234 | 73,85397 | 96,69755 | AgriculturaiRamaderia |
| 166,6873 | 172,4016 | 168,6269 | 0,919688 | 0,901840986 | 180,0134 | 207,8616 | Transport |
| 154,2152 | 164,2152 | 156,6398 | 0,577287 | 0,418769178 | 98,49734 | 120,6341 | Serveis |
| 156,3574 | 173,1574 | 159,2668 | 0,906426 | 0,852955793 | 99,08242 | 129,7293 | Residus |
| 188,5941 | 198,5941 | 191,0186 | 0,940138 | 0,917689374 | 412,6007 | 505,3306 | Total |

Comparison of Model Indices

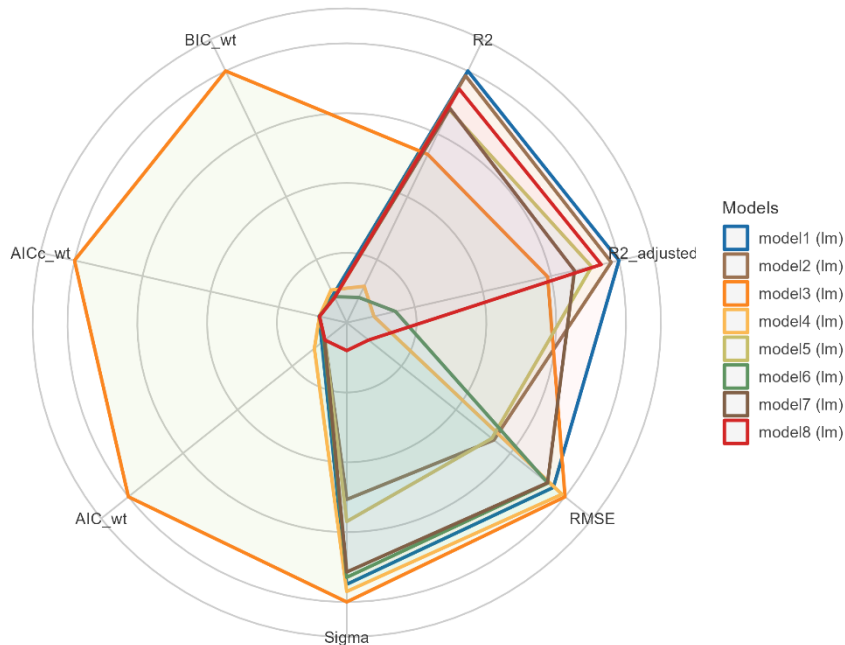


Figura 2 Comparativa entre els models corresponents als diversos sectors i total.

Diagnòstic dels models

També s'han realitzat comprovacions estadístiques per verificar les hipòtesis dels models lineals:

- normalitat dels residus,
- homoscedasticitat,
- absència de multicol·linearitat entre predictors.

La multicol·linearitat s'ha analitzat mitjançant el càlcul dels factors d'inflació de variància (VIF). Aquest procés permet detectar possibles redundàncies entre variables explicatives i assegurar l'estabilitat dels coeficients estimats.



Figura 3 Exemple del resum gràfic de les principals avaluacions realitzades per a la variable emissions totals. S'han portat a terme avaluacions similars per a cadascuna de les variables estudiades.



Validació retrospectiva (backtesting)

Per avaluar la capacitat predictiva dels models, s'ha implementat un procediment de validació retrospectiva (*backtesting*).

Aquest procés consisteix a:

1. eliminar els últims anys de la sèrie històrica,
2. ajustar el model únicament amb les dades anteriors,
3. predir els anys exclosos,
4. comparar les prediccions amb els valors realment observats.

La validació s'ha realitzat considerant diferents longituds de període de test (3, 5 i 7 anys), amb l'objectiu d'analitzar l'estabilitat temporal dels models.

Els errors de predicció s'han quantificat mitjançant:

- RMSE,
- error absolut mitjà (MAE),
- errors relatius percentuals,
- biaix mitjà.

Aquest procediment permet obtenir una estimació més robusta de la capacitat predictiva real dels models fora de mostra.

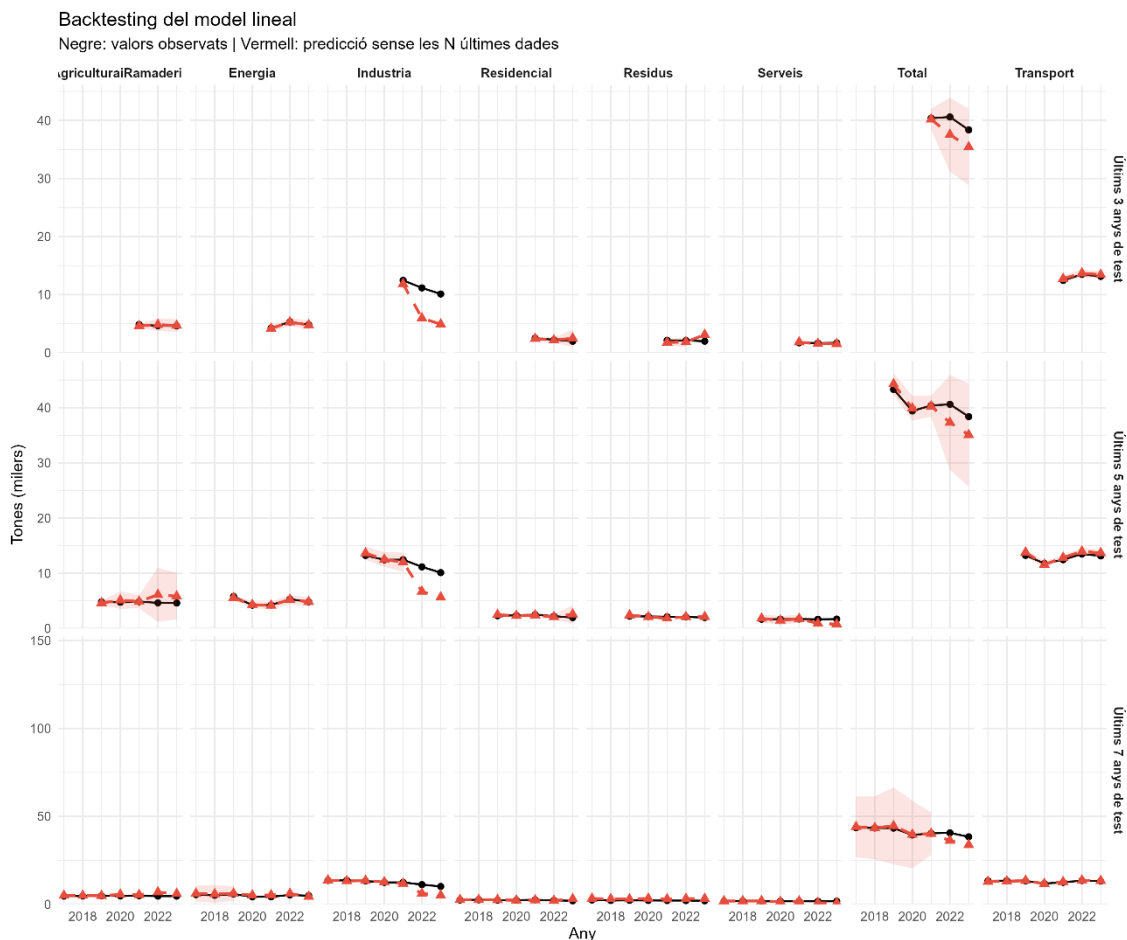


Figura 4 Resum gràfic dels backtesting realitzats per a les diferents variables.

Generació d'escenaris futurs

Un cop validats els models, s'han utilitzat per generar prediccions futures sota múltiples escenaris. Per a cada escenari i any projectat, el model calcula:

- la predicció central de les emissions,
- i els corresponents intervals de predicció.

Els intervals de predicció permeten representar explícitament la incertesa associada a les estimacions futures.

Finalment, els resultats s'han representat gràficament per facilitar-ne la interpretació. Els gràfics inclouen:

- comparacions entre valors observats i predits,
- distribució dels residus,
- validacions de backtesting,
- i evolució temporal de les emissions sota diferents escenaris.

Tots els resultats i mètriques s'han exportat automàticament a fitxers Excel per facilitar-ne l'anàlisi i documentació posterior.

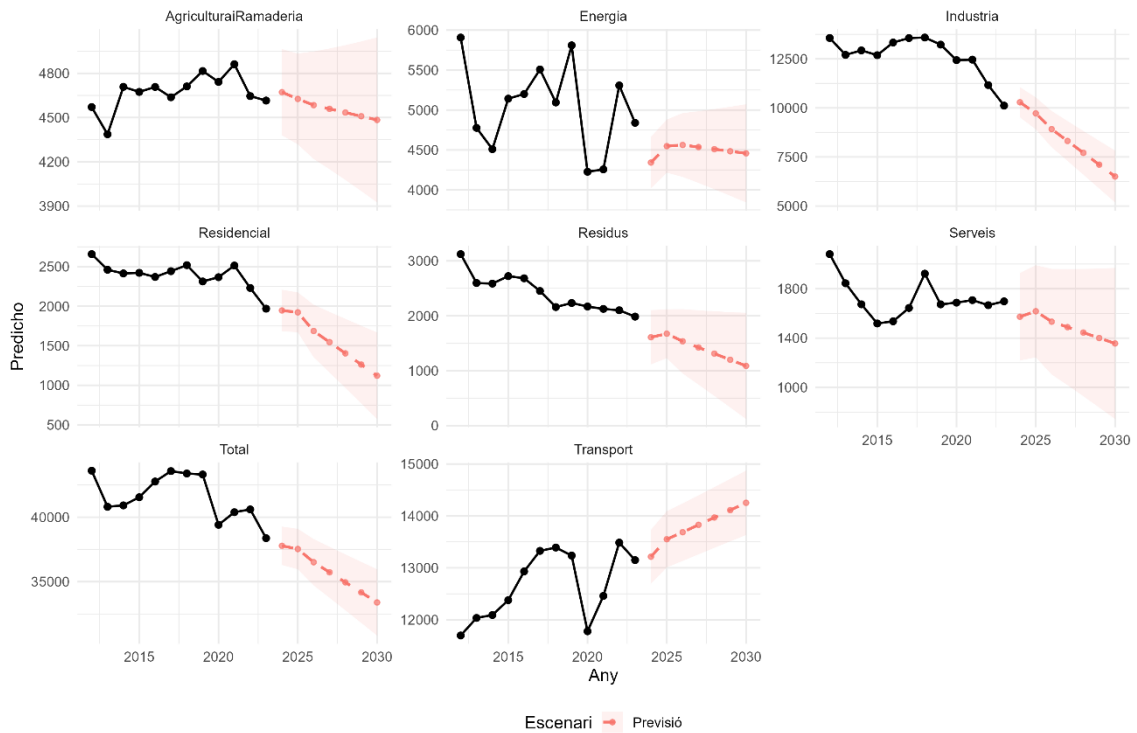


Figura 5 Representació gràfica dels resultats obtinguts amb els corresponents intervals de predicció.