

Big Data i Intel·ligència Artificial per a la prevenció d'epidèmies



Monitoratge i predicció per
a la detecció primerenca de
brots epidèmics



Socis del projecte

Entitats públiques i sanitàries



Centres de R+D+i



Universitats





Autoria	7
1. Resum Executiu.....	11
2. Motivació, Context i Objectius.....	17
3. Fonts de dades.....	19
Epidemiològiques	20
Mobilitat	23
Ambientals	24
4. Arquitectura Tecnològica i Metodologia	25
Descripció general de sistema	26
Capes geogràfiques	27
Accés a les dades	28
Descripció dels Models Epidemiològics	29
5. El Model Epidemiològic	31
Model Ambiental	32
Model de Mobilitat	33
Model Epidemiològic Mecànic.....	33
Model Epidemiològic Dirigit per Dades	35
6. Anàlisi	37
Estudi de la mobilitat i els determinants socioeconòmics	38
Simulacions d'escenaris pandèmics amb el model epidemiològic mecànic.....	51
Descripció i resultats del calibrat.....	53
Simulacions d'escenaris pandèmics amb el model dirigit per dades	58
Reflexions i recomanacions sobre les fonts de dades	61
7. Conclusions i recomanacions finals	63

8. Annexos67**Annex I: Origen de les dades del model de mobilitat.....68****Annex II: Model Ambiental en detall.....72**

Població i període d'estudi	72
Variables resultat.....	72
Variables meteorològiques.....	72
Contaminants atmosfèrics	72
Estimació dels nivells de les variables meteorològiques i dels contaminants atmosfèrics.....	72
Variables socioeconòmiques	73
Models d'associació	73
Resultats	74
Conclusions	75
Referències	76

Annex III: Model Epidemiològic amb Equacions Diferencials en detall93

1 COVID-19 and input data	93
2 Mobility model	97
3 Epidemic model.....	98
4 Numerical simulation of the Covid-19 spreading	107
5 Calibration/Parameter inference.....	111
References	117

Autoria



Autor	Centre	Contribució
Dr. Vicent Ribas Ripoll	Eurecat Data Analytics in Medicine	Investigador Principal
Dr. Felip Miralles Barrachina	Eurecat Digital Health	Coordinador
Dra. Odei Rey Orozco	Eurecat Data Analytics in Medicine	Model Epidèmic Mecànic
Xavier Rafael Palou	Eurecat Data Analytics in Medicine	Model Dirigit per Dades
Paula Subías Beltran	Eurecat Data Analytics in Medicine	Model Epidèmic Mecànic
Dr. Marc Torrent	Eurecat Data Science and Big Data Analytics	Estratègia i Models de Mobilitat
Dr. Julià A. Vicens Bennassar	Eurecat Data Science and Big Data Analytics	Models de Mobilitat
Dr. Marc Saez	Grup de Recerca en Estadística, Econometria i Salut (GRECS), Universitat de Girona CIBER de Epidemiología y Salud Pública (CIBERESP)	Model Mediambiental
Dra. Maria Antònia Barceló	Grup de Recerca en Estadística, Econometria i Salut (GRECS), Universitat de Girona CIBER de Epidemiología y Salud Pública (CIBERESP)	Model Mediambiental
Dr. Miguel Ponce-de-Leon	Barcelona Supercomputing Center (BSC-CNS)	Arquitectura Tecnològica i Metodologia
Dr. Alfonso Valencia	Barcelona Supercomputing Center (BSC-CNS) - ICREA	Arquitectura Tecnològica i Metodologia
Dr. Alex Arenas	Universitat Rovira i Virgili	Models epidemiològics i validació
Dra. Pilar Saura	Departament de Salut de la Generalitat de Catalunya	Requeriments i validació de resultats

Taula 1. Autoria

Revisor	Centre
David Ferrer	Secretari de Polítiques Digitals de la Generalitat de Catalunya
Daniel Marco	Director general d'Innovació i Economia Digital de la Generalitat de Catalunya
Pilar Saura	Directora General de Planificació en Salut de la Generalitat de Catalunya
Robert Fabregat	Director General de Recerca i Innovació en Salut de la Generalitat de Catalunya
Joan Mas	Director del CIDAI
Felip Miralles	Director de la Unitat de Tecnologia de Digital Health d'Eurecat
Vicent Ribas	Cap de línia d'Analítica de Dades en Medicina de la Unitat de Digital Health d'Eurecat
Myriam Garcia Berro	Directora de Tecnologia i Transferència d'Eurecat
Josep Maria Martorell	Director Associat del BSC
Alfonso Valencia	Director del Departament de ciències de la vida del BSC
Antoni Plasencia	Director General ISGlobal
Alexandre Arenas	Professor i investigador de la Universitat Rovira i Virgili
Marc Saez	Professor i investigador de la Universitat de Girona
Maria Antonia Barceló	Professora i investigadora de la Universitat de Girona
Ventura Clotet	President Fundació Lluita contra la Sida i les Malalties Infeccioses
Oriol Mitjà	Investigador de la Fundació Lluita contra la Sida i les Malalties Infeccioses
Antoni Trilla	Cap de Servei de Medicina Preventiva i Epidemiologia de l'Hospital Clínic

Taula 2. Revisió



Resum Executiu



El projecte Big Data i Intel·ligència Artificial per a la prevenció d'epidèmies va néixer durant la primera onada de la pandèmia de la COVID-19, amb l'objectiu d'estudiar-ne l'evolució i plantejar diferents escenaris epidèmics que puguin servir per a mitigar la seva propagació. El projecte ha comptat des del seu inici amb el suport de les Conselleries de Polítiques Digitals i de Salut de la Generalitat de Catalunya.

Durant aquesta primera fase del projecte, l'equip s'ha enfocat principalment a identificar les fonts de dades més rellevants per a poder fer un seguiment el més acurat possible de la pandèmia. Per aquest motiu, s'ha considerat que qualsevol model que pugui resultar en prediccions fiables hauria de considerar els següents aspectes fonamentals: poder realitzar prediccions a nivell d'àrea bàsica de salut i prendre en consideració els patrons de mobilitat de les persones. Així mateix, també s'ha considerat adient estudiar l'impacte dels factors mediambientals en la transmissió de la malaltia.

Atenent a aquests requisits bàsics, les principals fonts de dades de salut que s'han emprat en tots els calibrats de models provenen de dades de l'INE i del programa PA-DRIS d'AQUAS les quals són les més acurades i precises. Una limitació d'aquesta font de dades és el seu retard en actualitzar-se a conseqüència dels procediments d'informació i confirmació de casos així com de facturació. Per tant, com que l'objectiu és realitzar prediccions sobre l'evolució pandèmica a Catalunya, s'han emprat les dades obertes que ofereix el web Govern Obert de la Generalitat¹. Les dades mediambientals també s'han extret del mateix portal en la seva major part. També es van utilitzar les dades diàries de les 29 estacions automàtiques de la 'Agència Estatal de Meteorologia' (AEMET), que va cedir-les, al Grup de Recerca en Estadística, Econometria i Salut (GRECS), Universitat de Girona.

Pel que fa a les dades de mobilitat, en un principi, es va assumir que es disposaria de dades d'operadors de telefonia mòbil però les dificultats alienes al projecte han fet impossible disposar d'aquesta valuosa font de dades. Això ha suposat un entrebanc en el desenvolupament

del projecte, que ha quedat palès en una desviació d'un mes en l'obtenció de resultats. Aquesta situació s'ha pogut redreçar la situació accedint inicialment a dades de mobilitat de l'INE i, posteriorment, del Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA).

S'havia previst que durant aquesta primera fase només es desenvoluparia un nou model mecànic basat en equacions diferencials seguint una metodologia del tipus susceptibles, exposats, infectats i recuperats (SEIR) que hauria de permetre interpretar els diferents elements implicats en la transmissió de la pandèmia (per exemple la taxa d'infecció, el nombre d'asimptomàtics, l'evolució dels ingressos en UCI i la mortalitat). Així mateix, aquest model ha de permetre interpretar els resultats obtinguts per un model dirigit per dades desenvolupat amb una metodologia d'aprenentatge profund (Deep Learning), que s'havia de desenvolupar en una fase posterior del projecte. Malauradament, la greu evolució de la pandèmia i l'inici de la tercera onada ha urgint a iniciar el seu desenvolupament ja en aquesta primera fase a fi i efecte de poder realitzar prediccions durant la tercera onada el més ràpid possible i tan bon punt s'actualitzen les dades al portal de dades obertes de la Generalitat.

Durant aquesta primera fase s'ha pogut desenvolupar el primer nou model mecànic i calibrar-lo. Aquest model compta amb vuit compartiments: susceptibles, asimptomàtics, infectats, recuperats, hospitalitzats, hospitalitzats en unitats de cures intensives, defuncions i recuperats. Les transicions entre els grups de susceptibles i asimptomàtics s'han estudiat en una primera implementació amb un model de mobilitat a partir de les dades de l'INE. També s'ha pogut realitzar un primer calibrat d'aquest model de la primera onada amb un resultat acceptable (6.53% d'error pel millor ajust sobre la mortalitat).

També s'ha dut a terme un anàlisi dels factors ambientals i socioeconòmics i, més concretament, quina és la seva relació amb la mobilitat, element clau en la propagació de l'epidèmia.

1 http://governobert.gencat.cat/ca/dades_obertes/

El model dirigit per dades s'ha desenvolupat emprant un model LSTM (Long short-term memory), que és un arquitectura de xarxa neuronal recurrent utilitzada en aprenentatge profund (Deep Learning). Aquest model s'ha entrenat amb les dades de la primera i segona onada de la pandèmia i s'ha testejat amb les dades de la tercera onada amb les dades epidemiològiques del portal de dades obertes i de mobilitat del MITMA integrades a l'eina flowmaps desenvolupada al BSC. A dia d'avui, aquest model presenta un error del 2% en la predicció de la

incidència acumulada. Totes les prediccions de l'evolució de la pandèmia s'han realitzat amb aquest model.

La següent gràfica mostra l'evolució diària dels casos de COVID-19 a tot el principat en el període 02/02/2020 a 22/01/2021. La corba taronja correspon als casos reals, la blava als casos predits pel model dirigit per dades i les zones vermelles corresponen al test realitzat durant la tercera onada incloent-ne els últims dies.

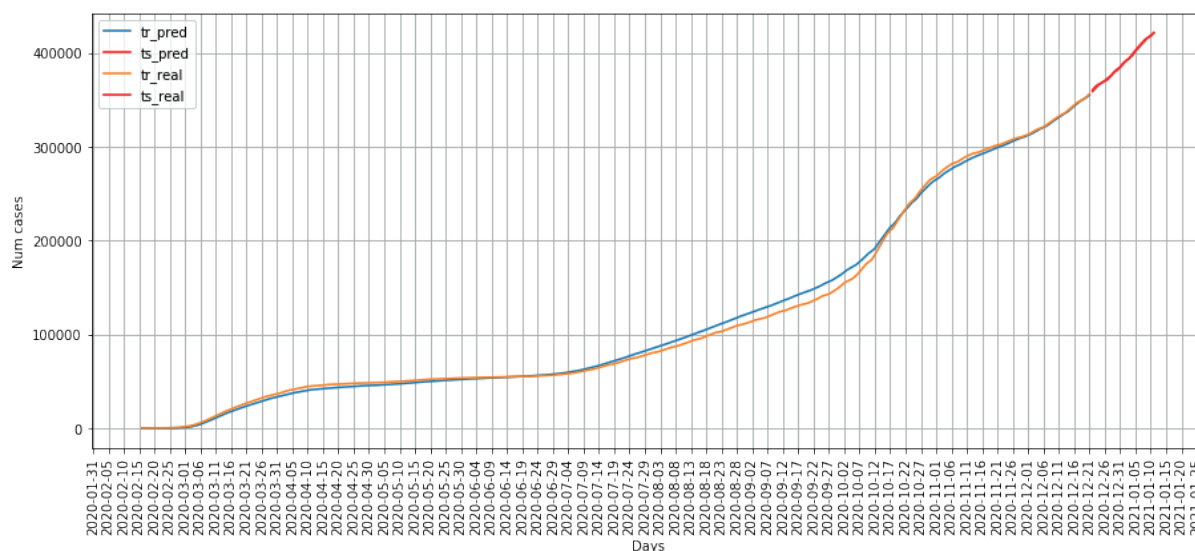


Figura 1. Evolució diària dels casos de COVID-19

L'error de predicció obtingut per la població de test és de 1531 casos. Independentment de l'error, la tendència de la corba és clarament creixent. Per aquest motiu, es va realitzar una predicció de l'evolució de la pandèmia prenent com a data d'inici el dia 18 de gener de 2021 (últim dia de dades disponible) fins el dia 7 de Febrer de 2021. Així mateix, també s'ha estudiat l' R_t (taxa d'infecció) durant aquest període. El següent gràfic mostra els resultats obtinguts de la predicció diària amb dades fins el dia 18/01/2021.

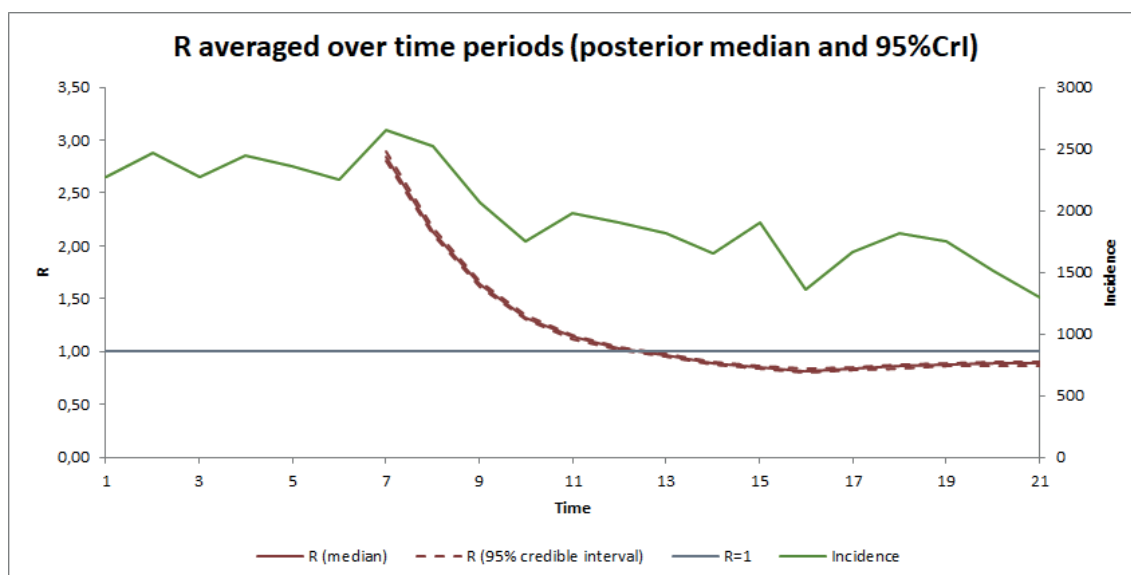


Figura 2. Rt (taxa d'infecció) diària

Tot seguit afegim l'última predicció diària fins el dia 12/02/2021 realitzada amb l'actualització de dades fins el dia 22/01/2021.

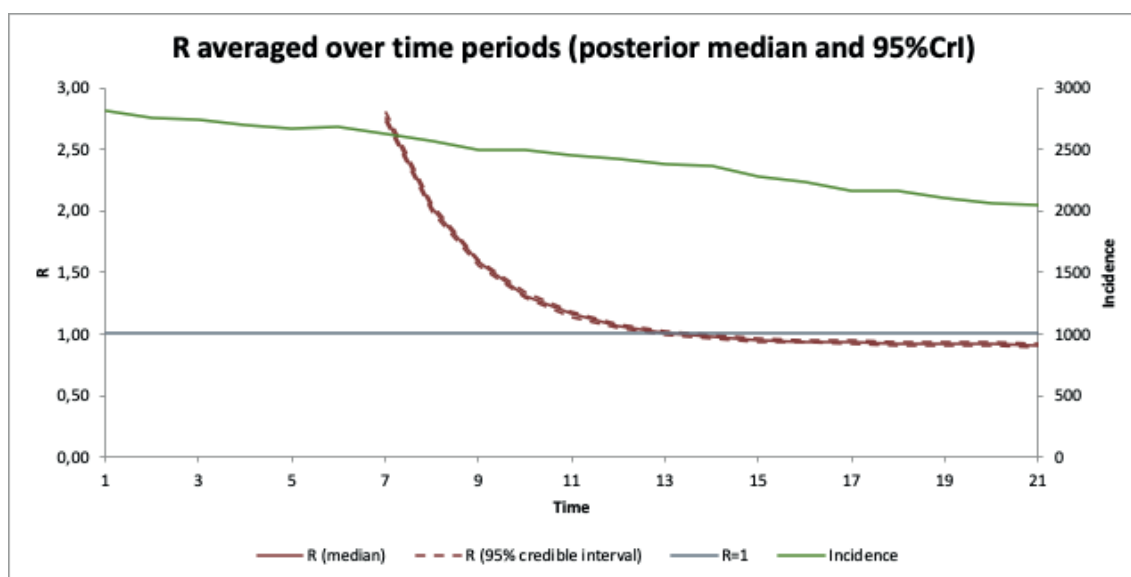


Figura 3. Rt (taxa d'infecció) diària

Malgrat que la tendència del nombre de casos i de l' R_t (taxa d'infecció) sigui a la baixa s'observa que aquest darrer paràmetre s'estabilitza a 0,94. La interpretació d'aquest resultat és que la pandèmia seguirà amb un creixement exponencial fins a finals de Gener i s'estabilitzarà en un creixement lineal durant la primera setmana de Febrer. Aquest comportament es replica amb l'ocupació de llits de cures intensives que es presenta tot seguit (i.e. creixement exponencial fins al final de Gener i lineal durant Febrer). Cal comentar també que el model actual no pren en consideració la propagació de la soca britànica B117 al principat, que a data d'aquest informe s'estima que afecta un 5% dels infectats. Tot i això, l'augment que observem en la R_t estimada podria estar alineat amb

l'augment de la taxa d'infecció que s'observa amb la soca britànica. L'augment sostingut d'aquesta soca implica un augment de la R_t en el temps podria acabar suposant un creixement exponencial en un horitzó temporal de tres setmanes. Aquesta evolució es monitoritzarà en les futures actualitzacions que es realitzin de les prediccions a mesura que les dades es vagin actualitzant.

La tendència actual en l' R_t calculada es correspon amb la previsió de creixement calculat en el nombre acumulat d'ingressos en UCI. A la següent gràfica mostrem es mostra la previsió dels ingressos acumulats en UCI fins el 12 de Febrer.

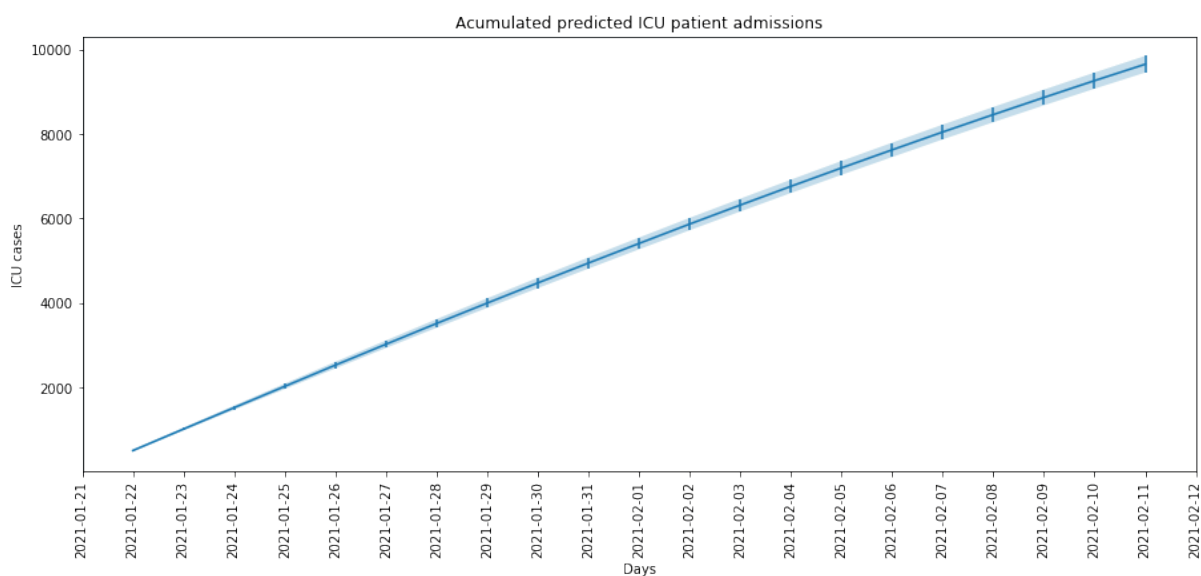


Figura 4. Previsió dels ingressos acumulats a l'UCI

Finalment la següent gràfica mostra la previsió diària d'ingressos a l'UCI. La mitjana prevista de casos diaris és de 447,6 +- 20,2 casos.

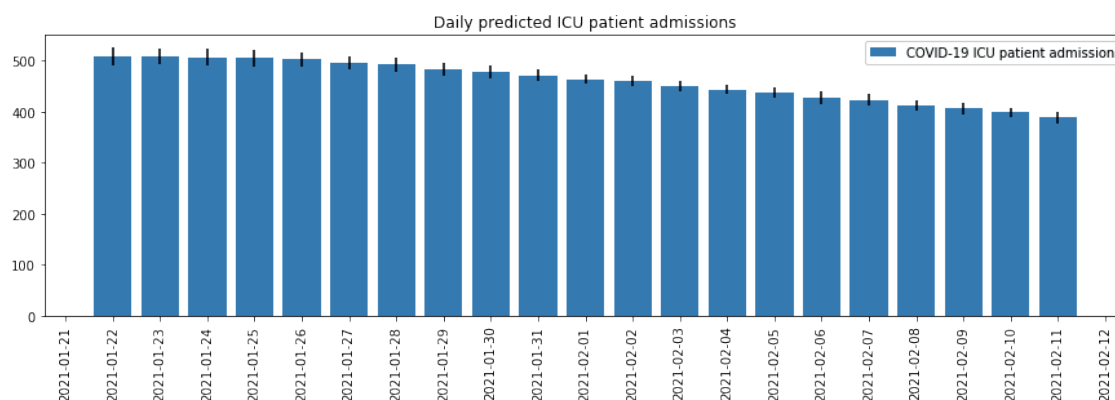


Figura 5. Previsió diària d'ingressos a l'UCI

Els passos següents, a plantejar en una segona fase del projecte, consisteixen en actualitzar les prediccions realitzades a mesura que es vagi disposant d'actualitzacions a les dades, completar el sistema de predicció d'ingressos hospitalaris i en unitats de cures intensives, millorar tots els models desenvolupats en aquesta primera fase i proporcionar un quadre de comandament i un sistema d'alertes per als usuaris finals.

La propagació del SARS-CoV-2 a tot el món s'ha convertit en una preocupació global amb la proliferació de recomanacions a nivell nacional i local que canvien diàriament a mesura que es disposa de nova informació i pronòstics obtinguts de diferents models i de la pròpia evolució de la pandèmia. La comprensió de la transmissió del virus tant a nivell demogràfic com geogràfic és fonamental per a la planificació adequada d'intervencions de salut pública a partir de la valoració de diferents escenaris possibles obtinguts a partir de diferents models matemàtics.

Motivació, Context i Objectius



Atenent a la natura complexa i multifactorial de les pandèmies (p.e. grip A, MERS-CoV o COVID-19) es fa necessari evolucionar l'enfoc epidemiològic cap a marcs conceptuals de salut multinivell els quals permeten integrar i estudiar múltiples factors de risc. Aquest model multinivell no només ha de caracteritzar cada nivell sinó focalitzar-se principalment en 2 àrees principals: i) l'entorn de l'individu com, per exemple, la seva àrea de residència, la seva mobilitat i exposicions climatològiques (temperatura i humitat), ii) el sanitari.

Està àmpliament acceptat per la comunitat científica que la gran disponibilitat de dades provinents de múltiples fonts d'informació pel seu posterior anàlisi, pot beneficiar significativament els models epidemiològics existents. Els avenços de les tecnologies Big-Data han permès la creació de grans repositoris de dades en epidemiologia i mobilitat. Per exemple, les matrius de mobilitat i dades de geolocalització poden utilitzar-se per a estudiar la propagació d'una malaltia a nivell poblacional, per a estimar taxes de reproducció i, per tant, alimentar models d'infecció per a provar hipòtesis i escenaris relacionats amb les interaccions entre persones, exposicions, propagació i pronòstic de la malaltia a fi i efecte de poder prendre les millors decisions possibles en un context determinat.

El principal objectiu d'aquest projecte és desenvolupar un model matemàtic mecanicista que descriu el procés epidemiològic a cada una de les diferents àrees bàsiques de salut (ABS), utilitzant dades epidemiològiques, i incorporant dades de mobilitat, dades ambientals i dades demogràfiques. Així mateix, també s'ha desenvolupat un model dirigit per dades desenvolupat amb tècniques d'aprenentatge automàtic tota la seqüència de dades epidèmiques de que es disposa al portal de dades obertes de la Generalitat i també amb les dades de mobilitat del MITMA.

L'anàlisi dut a terme permet estudiar diferents estadis del procés epidèmic i posar en crisi el model per així poder analitzar com impacten determinades mesures a les diferents variables epidemiològiques (el nombre de morts, l'ocupació hospitalària, l'ocupació a unitats de cures intensives, etc.). A més a més, s'ha dut a terme una anàlisi dels factors ambientals i socioeconòmics i més concretament quina és la seva relació la mobilitat, element clau en la propagació de l'epidèmia.

Fonts de dades



Com ja s'ha comentat anteriorment, l'objectiu principal d'aquesta primera fase del projecte ha consistit en l'elaboració d'un primer model matemàtic, que conté dades epidemiològiques agregades, matrius de mobilitat entre Àrees Bàsiques de Salut (ABS) i dades mediambientals per l'estudi de l'evolució de la pandèmia de la COVID-19.

Epidemiològiques

Pel que fa a les dades epidemiològiques, s'han contemplat diverses fonts de dades per a:

- L'estudi de l'evolució temporal de les infeccions per a cada ABS.
- L'estudi de l'evolució d'ingressos hospitalaris i a les unitats de cures intensives per a cada ABS.
- L'estudi de l'evolució de la mortalitat agregada per a cada ABS.

De forma més específica, durant aquesta primera fase del projecte s'ha estudiat les dades de les següents fonts:

Origen	Descripció	URL
Instituto Nacional de Estadística	Defuncions causades per COVID-19	https://www.ine.es/jaxi/Datos.htm?path=/t15/p417/covid/&file=01010.px#!tabs-grafico
Instituto de Salud Carlos III – MoMo	Sistema de monitorització de la mortalitat diària	https://momo.isciii.es/public/momo/dashboard/momo_dashboard.html
Dades obertes de Catalunya	Dades segregades per sexe i municipi	https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/jj6z-iyrp
Dades obertes de Catalunya	Dades segregades per sexe i ABS	https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/xuwf-dxjd
Dades obertes de Catalunya	Registre central de població del CatSalut	https://analisi.transparenciacatalunya.cat/Salut/Registre-central-de-poblaci-del-CatSalut/ftq4-h9vk
Dades obertes de Catalunya	Registre de defuncions per COVID-19 a Catalunya	https://analisi.transparenciacatalunya.cat/Salut/Registre-de-defuncions-per-COVID-19-a-Catalunya-Se/uqk7-bf9s
Dades obertes de Catalunya	Evolució dels casos, ingressats i defuncions a Catalunya	https://app.powerbi.com/view?r=eyJrljoiO-DA4Njl3ZmItYjgyYS00MTUzLTlkNWEtMD-MyYzg0YzA0MmY3liwidCl6ljNiOTQyN2RjLW-QzMGUtNDNiYy04YzA2LWZmNzI1MzY3Nm-ZlYyYlslmMiOjh9&pageName=ReportSection-882f295efba528bcbcad

Taula 2. Fonts de dades epidemiològiques

Així mateix, també s'ha accedit a les dades detallades del programa PADRIS d'AQUAS. Aquest conjunt de dades és el que s'ha utilitzat per a la calibració del model epidemiològic. Aquestes dades han estat agregades i anonimitzades complint tant el protocol d'AQUAS sobre privacitat com la normativa RGPD. Amb la finalitat d'extraure les dades s'han identificat els següents codis ICD-10, que identifiquen les infeccions per SARS-COV2:

1. U07.1 COVID-19 (virus identificat)
2. U07.2 COVID-19 (sosпита d'infecció per COVID-19 durant la primera onada).

S'ha separat les dades per ABS. Així mateix, també s'ha estratificat la població en tres rangs d'edat: [0, 25), [25, 50) i ≥ 50 anys. S'han detallat els següents camps per a cada registre:

Per a cada ABS:

- Rang d'edat.
- Data primera PCR positiva.
- Nombre de PCR positives.

I per a cada ABS, grup d'edat, i data primera PCR positiva

- Nombre d'hospitalitzacions: si una persona hospitalitza dues vegades, compta com a dues hospitalitzacions.
- Mitjana de dies de hospitalització.
- Nombre d'ingressos en UCI.
- Mitjana de dies d'ingrés en UCI, excloent els dies d'hospitalització.
- Mortalitat hospitalària.
- Mortalitat no hospitalària a menys de 30 dies.
- Mortalitat no hospitalària a 30 i 60 dies.
- Mortalitat no hospitalària a més de 60 dies.

L'extracció d'aquestes dades inclou l'afectació per COVID-19 del 28-02-2020 al 31-08-2020. Hi ha un total de 48.786 observacions que descriuen les dinàmiques de les 373 ABS de Catalunya.

A la següent figura veiem les dues primeres onades de la pandèmia i identifiquem de forma clara les setmanes de confinament domiciliari.

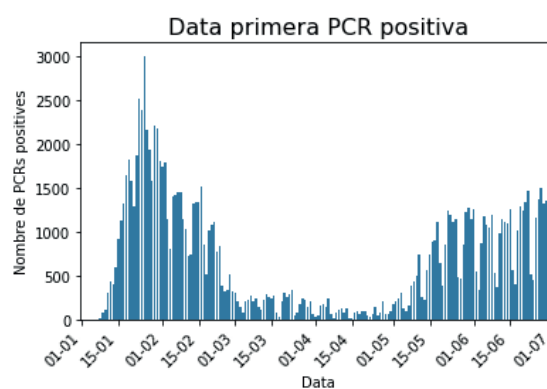


Figura 6. Nombre de PCRs positives a les dues primeres onades de la pandèmia

En la següent figura es mostra el número de PCRs positives per cada grup d'edat en el temps. El grup més afectat és el format per persones majors de 50 anys.

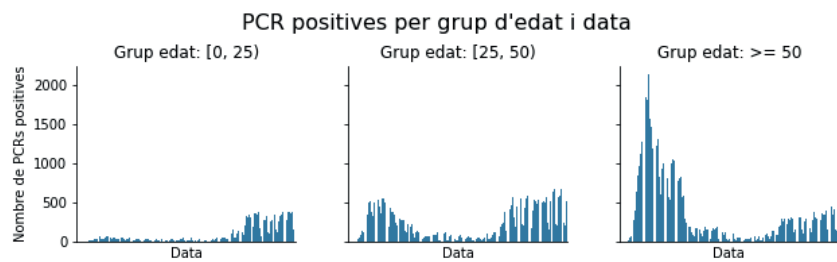


Figura 7. Número de PCRs positives per cada grup d'edat

A continuació es mostren dues gràfiques que representen la mitjana de dies d'hospitalització (esquerra) i ingrés a la UCI (dreta). Es pot notar que, en mitjana, els pacients de la COVID-19 van passar més temps a la UCI que hospitalitzats.

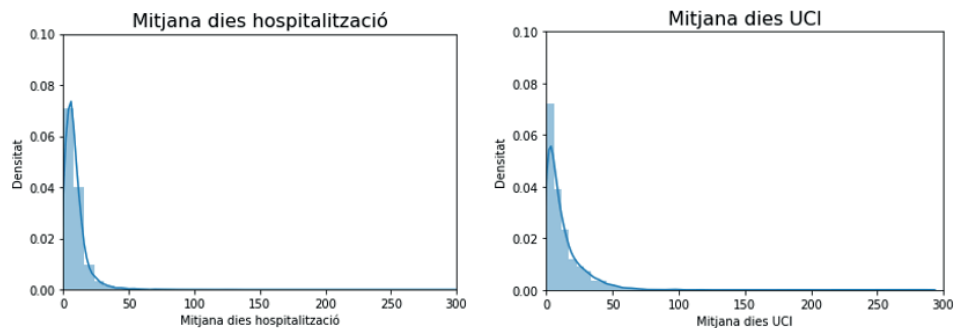


Figura 8. Mitjana de dies d'hospitalització (esquerra) i ingrés a l'UCI (dreta)

Per últim, trobem de rellevància mostrar la següent gràfica, que mostra que a major edat, més llarga va ser, en mitjana, l'estància a la UCI.

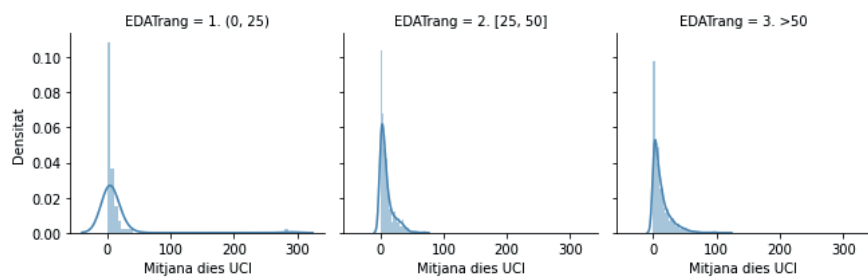


Figura 9. L'estància a l'UCI per cada grup d'edat

Mobilitat

L'estudi de la mobilitat es duu a terme utilitzant com a font de dades principal les dades de mobilitat extretes a partir de registres anonimitzats de telefonia mòbil. Les dades, processades i facilitades pel Ministerio de Transporte, Movilidad y Agenda Urbana (MITMA) del Gobierno de España, corresponen al període que va des de 14 de febrer fins al 31 de desembre, i es continuen actualitzant. L'abast territorial de les dades inclou tot Espanya, encara que en aquest estudi ens centrem en els viatges que tenen origen o destí Catalunya. El territori està dividit en àrees de mobilitat que vénen definides per: districtes, municipis, agrupació de districtes i agrupació de municipis segons la densitat de població, en total n'hi ha al voltant de 450 a tot Catalunya.

Concretament el MITMA proporciona dues matrius amb la informació de mobilitat. La primera és una matriu diària amb els fluxos de viatges entre diferents àrees de mobilitat que indica el nombre total de viatges entre àrees per cada dia i hora amb metainformació d'aquest, com els motius de viatge a origen i destí (casa, feina, altres), i el total de kilòmetres dels viatges, entre d'altres. La segona, és una matriu diària amb el nombre de viatges que fa la població d'una àrea de mobilitat, concretament la població d'una determinada àrea de mobilitat que fa 0, 1, 2 o +2 viatges.

Pel cas que ens ocupa els viatges totals entre diferents àrees de mobilitat, així com els viatges dins la mateixa àrea es poden organitzar com una col·lecció de xarxes ponderades temporalment on cada àrea de mobilitat correspon a un node de la xarxa i cada enllaç entre nodes representa el nombre de viatges. Aquestes les utilitzem per analitzar com les mesures de salut pública, les restriccions i demés factors socials impacten en la mobilitat, així com pel càlcul dels paràmetres de mobilitat del model epidemiològic.

Les dades obertes i la informació metodològica es pot trobar en el següent enllaç:

<https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data>



Figura 10. Mobilitat diària entre diferents àrees.

A la figura a es mostra la mobilitat diària (núm. viatges) amb origen L i destí B el dia t (5 viatges), així com el total de viatges de B a L (3 viatges).

La figura b mostra la xarxa de mobilitat diària entre B, G i L.

La figura c) mostra la matriu origen-destí, una representació matricial de la xarxa de mobilitat.

Ambientals

Com a variables mediambientals es van utilitzar tant variables meteorològiques com contaminants atmosfèrics.

Les variables meteorològiques considerades van ser les mitjanes diàries de la temperatura mitjana (mesurada en graus centígrads, °C), la humitat relativa (mesurada com percentatge), la pressió atmosfèrica (mesurada en hecto Pascals, hPa), la velocitat del vent a 2m (mesurada en metres per segon, m/s) i la irradiació solar global (mesurada en Wats per metre quadrat, W/m²), des del 1 de gener fins el 31 d'agost de 2020, ambdós inclosos.

Dues van ser les fonts d'aquestes variables, per una banda es van obtenir del portal de dades obertes de la Generalitat de Catalunya (<https://analisi.transparenciacatalunya.cat/Medi-Ambient/Dades-meteorol-giques-de-la-XEMA/nzn-apee>), que conté les dades de les 188 estacions de la Xarxa d'Estacions Meteorològiques Automàtiques (XEMA) del Servei Meteorològic de Catalunya (METEOCAT) (Generalitat de Catalunya). Per altra banda, es van utilitzar les dades diàries de les 29 estacions automàtiques de la 'Agència Estatal de Meteorologia' (AEMET), qui va cedir-les, al Grup de Recerca en Estadística, Econometria i Salut (GRECS), Universitat de Girona.

Pel que fa als contaminants atmosfèrics, en els models es van utilitzar les mitjanes diàries del nivell de partícules gruixudes, amb un diàmetre de 10 micres (µm) o menys, PM₁₀, del nivell de diòxid de nitrogen, NO₂ del nivell d'ozó, O₃. També es van utilitzar, encara que no en els models, les mitjanes diàries dels nivells de les partícules fines, amb un diàmetre de 2,5 µm o menys, PM_{2.5}, d'òxid de nitrogen, NO, de diòxid de sofre, SO₂, de monòxid de carboni, CO, d'hidrocarburs aromàtics (benzè, C₆H₆ i benzopirè, B), i de metalls pesats (plom, arsènic, níquel i cadmi). En tots els casos el període va ser 1 de gener de 2011 fins el 31 d'agost de 2020.

Totes les dades relatives a la contaminació atmosfèrica també es van obtenir del portal de dades obertes de la Generalitat de Catalunya (<https://analisi.transparenciacatalunya.cat/Medi-Ambient/Qualitat-de-l-ai-re-als-punts-de-mesurament-autom-t/tasf-thgu>). En aquest cas conté les dades provinents dels 144 punts

de mesurament automàtics de la Xarxa de Vigilància i Previsió de la Contaminació Atmosfèrica (XPVCA) (Departament de Territori i Sostenibilitat, Generalitat de Catalunya).

En els models també es van utilitzar, com a variables de control, distintes variables socioeconòmiques. En concret: la renda mitjana per persona (en euros) - mitjana dels anys 2015, 2016 i 2017 (Font: Atles de Distribució de Renda de les Llars, Institut Nacional d'Estadística, INE, https://www.ine.es/en/experimental/atlas/exp_atlas_tab_en.htm, dades obertes); la taxa d'atur el 2011 (en percentatges) (Font: Cens Espanyol de Població i Habitatges 2011, INE, http://www.ine.es/censos2011_datos/cen11_datos_resultados_seccen.htm, dades obertes); percentatge de població de 65 anys o més el 2019 (Font: Estadística del Padró Continu, INE, https://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736177012&-menu=resultados&secc=1254736195461&i-dp=1254734710990#!tabs-1254736195557, dades obertes); percentatge d'estrangers el 2019 de països amb índex de desenvolupament humà mitjà i baix segons el Programa de les Nacions Unides per al Desenvolupament (PNUD, 2019)(Font: Atles de Distribució de Renda de les Llars, INE, https://www.ine.es/en/experimental/atlas/exp_atlas_tab_en.htm, dades obertes); infrahabitatge - percentatge d'habitatges amb menys de 45 m² de superfície habitable el 2011 (Font: Cens Espanyol de Població i Habitatges 2011, INE, http://www.ine.es/censos2011_datos/cen11_datos_resultados_seccen.htm, dades obertes); i densitat de població el 2019 (en habitants per km²) (Font: Estadística del Padró Continu, INE, https://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736177012&-menu=resultados&secc=1254736195461&i-dp=1254734710990#!tabs-1254736195557, dades obertes).

Per últim, també es va controlar per la mobilitat de les persones. En concret, es va utilitzar el percentatge de persones que es mouen en l'ABS i el percentatge de persones que es mouen fora de la mateixa àrea, respecte a les que es mouen en l'ABS.

Arquitectura Tecnològica i Metodologia



Descripció general de sistema

S'ha dissenyat i desenvolupat un Sistema d'Informació Geogràfica basat en sistema de base de dades no relacional (MongoDB) per a la gestió de les dades de salut i mobilitat, dades geogràfiques i dades censals (Figura 11). S'ha dissenyat un model dades flexible que permet la incorporació de noves fonts de dades. Això ha estat una necessitat des de l'inici pel fet que el model ha de permetre adaptar-se a estructures de dades heterogènies que són definides pels diferents proveïdors. Per exemple, les diferents autoritats sanitàries que reporten dades a través d'un portal d'accés públic solen utilitzar formats diferents.

En termes més tècnics, el model de dades està orientat a documents i es troba organitzat en diverses col·leccions. Per exemple, la col·lecció *dades* s'usa per emmagatzemar totes les dades corresponents a sèries temporals de casos reportats de COVID-19 per municipi / zona hospitalària. Una altra col·lecció de dades clau inclou la col·lecció emprada per emmagatzemar les dades de mobilitat i la col·lecció emprada per emmagatzemar les diferents capes geogràfiques (Figura 11).

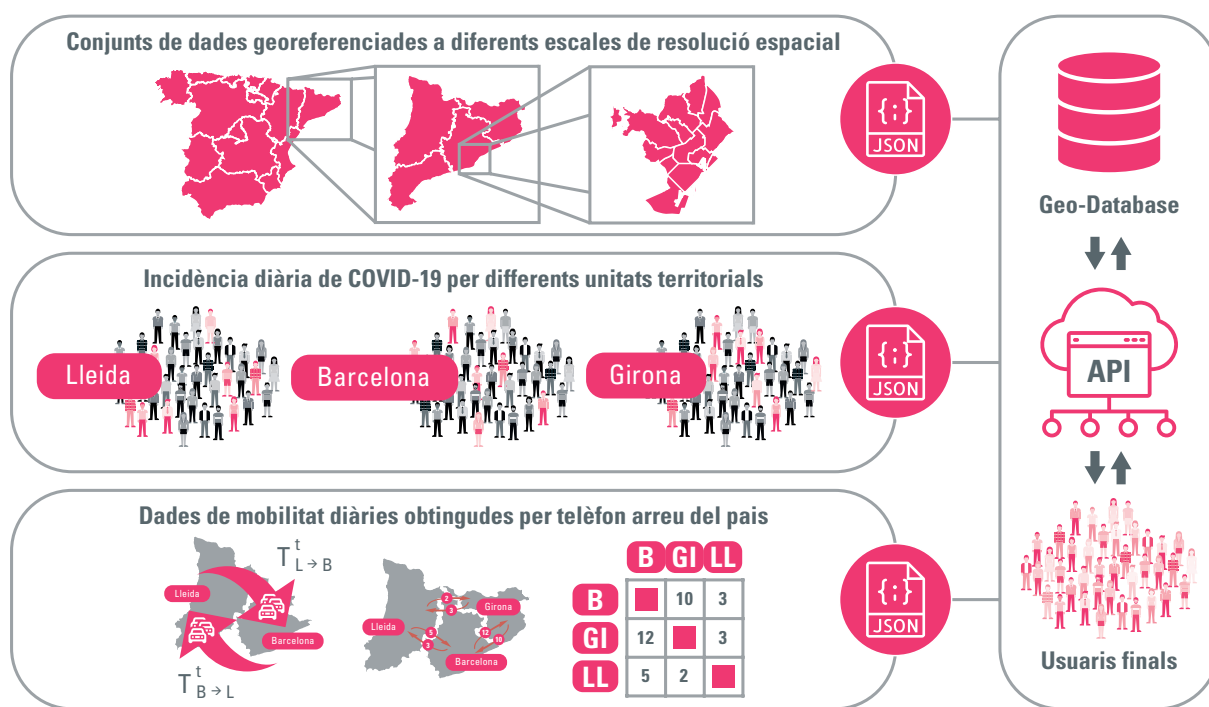


Figura 11. Esquema simplificat del Sistema d'Informació Geogràfica de Flow-Maps incloent les principals col·leccions de dades i la forma en què aquestes són exposades als usuaris finals.

Com es mostra a la Figura 11, les principals col·leccions de dades incloses en el sistema corresponen a informes de casos d'COVID-19 que s'informen diàriament en diferents nivells de resolució geogràfica i dades de mobilitat poblacional. Respecte a les dades de salut s'inclou els casos diaris a nivell de comunitat autònoma i províncies que reporta l'Institut de Salut Carlos III (ISCIII). D'altra

banda, hi ha informes diaris de casos COVID-19 a nivell d'ABS per a les vuit regions que ofereixen les dades: d'una banda Catalunya, Castella i Lleó, País Basc, Madrid, Navarra proveeixen dades a nivell d'ABS o zones bàsiques sanitàries (ZBS) i, per altra banda, Cantàbria, Astúries i València, reporten la incidència a nivell municipal.

Atès que les dades reportades per les diferents autoritats sanitàries no segueixen una estructura estàndard, s'ha definit un criteri comú per fusionar i consolidar les dades de fonts disperses que permet obtenir les sèries temporals per al menor nivell de resolució geogràfica possible. D'aquesta manera, per una banda sempre es conserven les entrades originals de dades tal com es reporten per l'autoritat sanitària corresponent i, al seu torn, processen les dades perquè s'emmagatzemen en un format estandarditzat. D'aquesta manera el sistema disposa d'una col·lecció de dades de salut consolidada que facilita la implementació de consultes genèriques. En el format consolidat, cada entrada té associada la font que ha reportat la dada, la data, un identificador de capa i polígon que permet geo-referenciar la dada i finalment el nombre de casos reportats en aquesta entrada. La descàrrega i consolidació es realitza de forma automàtica i periòdica.

Pel que fa a les dades de mobilitat, el sistema utilitza dades agregades i anonimitzades basades en telefonia mòbil que es proporcionen de manera oberta pel MITMA. Les dades reflecteixen els patrons de mobilitat diaris entre i dins de les zones de mobilitat, una capa geogràfica

definida segons la cobertura de les antenes de telefonia mòbil. Atès que les dades originals són informes per hora i aquesta granularitat no és necessària ni per a les anàlisis ni per als models, el sistema inclou una col·lecció de dades de mobilitat agregades per dia que inclouen les matrius origen-destí entre zones de mobilitat MITMA (veure Figura 12) per a cada dia. Anàlogament a les dades de salut, les dades de mobilitat s'actualitzen de forma periòdica, sempre que el proveïdor hagi reportat noves dades.

En el context de les bases de dades que integren diverses fonts externes, la procedència o *provenances* de la informació fa referència a la descripció de l'origen, història, i cicle de vida de les dades. Aquesta informació és fonamental, no només per a tasques de manteniment sinó per poder extreure l'origen de totes les fonts d'informació. Amb l'objectiu de proporcionar tota la informació necessària per identificar les fonts d'origen de les dades, així com el processament realitzat a la seva transferència/abocament, la base de dades compta amb una col·lecció que emmagatzema el *provenance* de totes les fonts de dades.

Capes geogràfiques

Com es descriu en la secció prèvia, les principals fonts de dades, com les dades d'incidència i mobilitat, es troben etiquetades amb identificadors relatius a una capa geogràfica. D'aquesta manera és possible mapejar les dades reportades (p.e., el nombre de casos positius confirmats en un municipi o a una ABS) a l'element geolocalitzat corresponent (i.e., al polígon geolocalitzat que correspon al municipi o a l'ABS). Com a conseqüència, la base de dades també inclou informació geogràfica de les unitats territorials en les quals es troben anotades les dades de salut i mobilitat així com altres capes d'utilitat.

En termes generals el sistema inclou les capes corresponents a les unitats territorials oficials que inclouen comunitats autònomes, províncies, municipis i districtes; en forma addicional s'han incorporat les capes corresponents a les zones hospitalàries, ZBS de 8 comunitats autònomes, així com la capa corresponent a les zones de mobilitat. Això permet realitzar consultes geogràfiques

integrades (p.e. nombre d'hospitals a una distància màxima de la zona amb major nombre de casos).

Les col·leccions de dades i capes es troben estructurades de manera uniforme per facilitar les consultes. En el cas de la col·lecció de capes, cada entrada o document conté les coordenades geogràfiques en ETRS89 (epsg:4258). Les principals capes es troben enumerades i descrites a la Taula 4 de l'Annex I. A la Figura 12 es representen les capes geogràfiques sobre les quals es troben anotades les dades de salut i mobilitat. Les dades d'incidència de casos que proveeix el ISCIII són a nivell de Comunitats autònomes i províncies (Figura 12a, b).

Per altra banda, les dades de mobilitat proveïdes pel MITMA estan geolocalitzades a la capa d'àrees de mobilitat, que són una combinació de districtes, municipis o agregacions dels anteriors en funció de la densitat poblacional (Figura 12c); en zones densament poblades.

com Madrid o Barcelona, les àrees de mobilitat solen correspondre amb districtes o grups de districtes, mentre que en àrees rurals menys poblades cada àrea de mobilitat pot correspondre amb un municipi o un conjunt de municipis. Per facilitar la integració de dades projectades

en diferents capes geogràfiques, com les dades de mobilitat i d'incidència, el sistema compta amb una col·lecció auxiliar en la qual s'han precalculat les relacions entre les diferents capes, ja que en molts casos no hi ha una correspondència exacta.

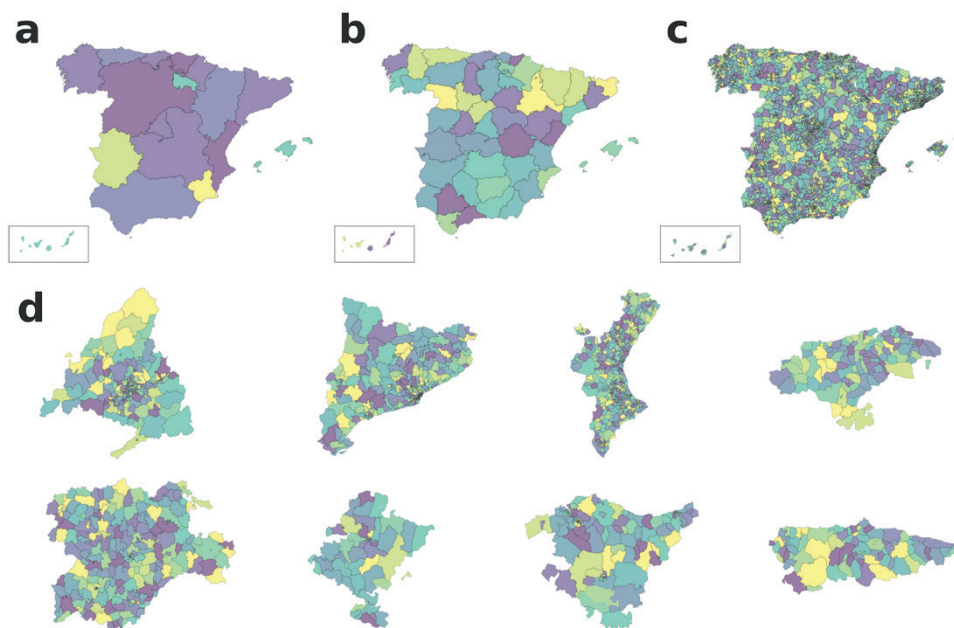


Figura 12. Capes geogràfiques en què estan anotades les diferents fonts de dades. En a i b es mostren les capes de comunitats autònomes i províncies, respectivament. En c es representen les àrees de mobilitat sobre les quals estan anotades les dades de mobilitat del MITMA.

En d es representen les capes en les quals es reporta incidència de casos de COVID-19 a major resolució espacial: d'esquerra a dreta de dalt a baix estan les comunitats de: Madrid (ZBS), Catalunya (ABS), València (municipis), Cantàbria (municipis), Castella i Lleó (ZBS), Navarra (municipis), País Basc (ZBS), Astúries (municipis).

Accés a les dades

Per brindar accés a les diferents col·leccions de dades s'ha dissenyat i implementat una interfície de programació d'Aplicacions (API, de l'anglès *Application Programming Interface*) que permet realitzar consultes de forma programàtica. L'objectiu de l'API és oferir un punt unificat d'accés a les dades que s'empri per consultar i extreure la informació necessària per calibrar models epidemiològics

desenvolupats per altres grups. L'API ha estat desenvolupada com un servei web emprant una arquitectura de tipus transferència d'estat representacional (REST, de l'anglès *REpresentational State Transfer*). D'aquesta manera l'API-REST pot també ser emprada com a punt d'accés d'aplicacions de tercers que vulguin consultar i extreure informació. La documentació de l'API juntament

amb diversos exemples de casos d'ús es troba accessible a través del següent enllaç: <https://flowmaps.life.bsc.es/api/apidocs>.

[es/api/apidocs](https://flowmaps.life.bsc.es/api/apidocs).

Descripció dels Models Epidemiològics

S'ha desenvolupat una metodologia per estudiar la dinàmica de la infecció per SARS-COV2. Aquesta metodologia ha de contemplar tres estadis principals:

- **Estadi latent:** aquest estat correspon al període de temps en el que un individu s'infecta i esdevé infecciós (i.e. pot transmetre el patògen a altres individus susceptibles).
- **Estadi infecciós:** és l'interval de temps durant el qual el hoste és infecciós (i.e. el patògen es transmet de forma directa o indirecta a altres individus). Aquest període es defineix com el temps entre el final de l'estadi latent el moment en el qual el hoste ja no transmet la infecció.
- **Estadi d'incubació:** és el lapse de temps entre l'exposició al patògen i el desenvolupament de símptomes o signes d'infecció.

En el cas particular de la COVID-19, els símptomes apareixen després de que l'individu esdevingui infecciós. En aquest escenari, el període latent té una durada inferior al període d'incubació, cosa que significa que els individus poden infectar-ne d'altres durant un temps sense mostrar cap símptoma notable. Aquesta etapa inicial o lleu d'infecció els símptomes de la qual es mantenen per sota del nivell de detecció clínica s'anomena infecció subclínica i la persona afectada s'anomena portadora asimptomàtica de la malaltia.

Aquest fet és especialment rellevant pel que fa a la transmissió de la malaltia i vindrà determinada en gran part per la mobilitat de les persones infectades asimptomàtiques. De fet, s'ha demostrat que la mobilitat de les persones en general està altament correlada amb la transmissió de la COVID-19 durant les tres onades de la pandèmia. Cal destacar també que el virus es pot transmetre des de la boca o el nas d'una persona infectada en petites partícules líquides quan tus, esternuda, parla, canta o respira

intensament. Aquestes partícules líquides tenen diferents mides, que van des de 'gotes respiratòries' més grans fins a 'aerosols' més petits. Altres persones poden infectar-se de COVID-19 quan el virus entra a la boca, al nas o als ulls, cosa que és més probable que passi quan les persones estan en contacte directe o proper (a menys d'un metre de distància) amb una persona infectada. L'evidència actual suggereix que la principal forma de propagació del virus és mitjançant gotes respiratòries entre les persones que estan en contacte estret. Així mateix, també s'ha demostrat recentment que la transmissió d'aerosols es pot produir en entorns específics, especialment en espais interiors, concorreguts i ventilats de manera inadequada, on les persones infectades passen llargs períodes de temps amb altres persones, com ara restaurants, pràctiques corals, classes de fitness, discoteques, oficines i / o llocs de culte. El virus també es pot propagar després que els infectats esternudin, tussin o toquin superfícies o objectes, com taules, poms i passamans.

En aquest context, també és d'especial rellevància contemplar l'impacte ambiental en la propagació de la pandèmia més enllà de la seva dependència amb la mobilitat sinó també en la transmissió i incidència de la malaltia a nivell comunitari.

A nivell metodològic, el model epidèmic que proposem està basat en un model compartimental, per a estudiar la dinàmica de la infecció de la forma més realista possible. En particular, es busca estudiar les transicions entre les sub-poblacions (per a cada ABS) de susceptibles, exposats, infectats/asimptomàtics, hospitalitzats, hospitalitzats en unitats de cures intensives, recuperats i morts. Així mateix, en aquest model, subsumeix la mobilitat com a mesura subrogada de la transmissibilitat i els factors mediambientals a les transicions entre susceptibles i exposats. Aquesta integració es realitza mitjançant un ajust de les equacions que modelen la força d'infecció al model.

Per altra banda, el model epidèmic proposat en aquesta primera fase del projecte està basat en un sistema d'equacions diferencials ordinàries pel qual s'ha realitzat un anàlisi d'existència i unicitat de solucions (i.e. s'ha comprovat que efectivament el sistema es pot solucionar) i d'estabilitat (i.e. que les solucions obtingudes són estables) mitjançant tècniques d'anàlisi en sistemes dinàmics.

Atenent a la pròpia dinàmica de la infecció, no s'ha pogut disposar d'una estimació fiable del nombre d'asimptomàtics. També, durant la primera onada de la pandèmia la confirmació diagnòstica per PCR va ser bastant limitada. Per aquest motiu, a nivell metodològic, es va decidir emprar les dades de mortalitat per a realitzar la calibració del model atenent a que és la dada més fiable de totes les que disposem. Finalment, la calibració del model s'ha realitzat mitjançant una aproximació Bayesiana amb un mètode de Monte Carlo, que s'ha implementat en un clúster de GPUs de 64GB, 6 nodes amb 384GB de RAM i 96 CPUs amb 100TB de disc dur.

El Model Epidemiològic



Model Ambiental

Aquest model s'alimenta de tres fonts de dades diferents, que són:

- **Dades epidemiològiques** proporcionades per AQUAS que contenen informació sobre la població catalana entre el 28/02/2020 i el 31/08/2020 (explicades de forma detallada a la secció Fonts de dades / Epidemiològiques). Les dades d'interès per al model ambiental són quatre: el nombre diari de casos positius per PCR per a aproximar la incidència de COVID-19, el nombre diari d'hospitalitzacions, el nombre diari d'ingressats a l'UCI, i el nombre diari de finats.
- **Dades meteorològiques.** S'han usat les variables esmentades a la secció Font de dades / Ambientals. Per tal d'estudiar les seves interaccions, s'ha dissenyat un índex de confort tèrmic.
- **Contaminants atmosfèrics.** S'ha avaluat l'exposició a llarg termini als contaminants esmentats a la secció Font de dades / Ambientals.

Les dades ambientals es recullen mitjançant estacions meteorològiques i punts de mesurament de la contaminació atmosfèrica. Aquestes instal·lacions no es distribueixen de forma homogènia per l'espai i la seva distribució tampoc correspon amb la de les ABS catalanes. En conseqüència, no sempre es pot establir una relació directa entre ABSs i estacions meteorològiques i punts de mesurament de la contaminació.

Per tal de solucionar aquest problema s'han modelitzat les variables meteorològiques i cadascun dels contaminants tenint en compte la dependència espacial i temporal, així com la superfície de les ABS. D'aquesta manera, s'aconsegueix una estimació dels valors de les variables ambientals per a cada ABS.

Conèixer la situació ambiental així com la incidència de COVID-19 a cada ABS, ens permet estudiar les relacions entre ambdós factors. Les variables meteorològiques d'interès són les següents: temperatura, humitat relativa, pressió atmosfèrica, velocitat del vent i irradiació solar, i els contaminants a estudiar: PM10, NO2 i O3.

El període d'estudi (febrer – agost) es separa en dos degut a que totes les variables d'interès mostren un comportament diferent en cada subperíode. Aquesta divisió crea els següents dos subperíodes: 28/02 - 30/06 i 1/07 - 31/08.

En els models s'estudia la relació entre les variables ambientals recent esmentades i altres variables que no només representen la incidència de COVID-19 sinó també la mobilitat i la demografia de les ABS. Donat que existeix una demora entre la infecció i la diagnosi i agreujament de la malaltia, tant les dades meteorològiques com les de mobilitat s'introdueixen amb una mitjana mòbil de 10 dies que permet estudiar els efectes d'aquestes variables 10 dies després de la seva recollida.

El model ambiental ens ha permet arribar a diferents conclusions, com que com més alta va ser la temperatura neta efectiva, major va ser el risc d'infecció. Aquest comportament també s'aprecia en el nombre d'hospitalitzacions i en el de morts: a major temperatura neta efectiva, major nombre d'hospitalitzacions i morts. De fet, a la segona onada es va apreciar una magnitud més gran que en la primera. En canvi, el nombre d'ingressos en UCI només es va veure afectat de forma significativa per la temperatura neta efectiva a la primera onada.

L'efecte de les variables meteorològiques va ser molt semblant al comportament de la mobilitat. A mesura que les condicions meteorològiques van ser més favorables, va haver-hi més mobilitat i per tant més contactes, el que va comportar un augment de la incidència.

L'exposició a la contaminació atmosfèrica, en particular per NO2, va estar associada a un augment de la incidència, encara que, sobre tot, durant la primera onada. Però, a més, aquesta associació només es va trobar en aquelles ABS més contaminades i en aquelles amb una contaminació moderada. Per tant, podria ser que l'exposició a la contaminació atmosfèrica fos un substitut d'altres variables que fossin les veritablement associades amb la incidència (baix nivell socioeconòmic, per exemple).

Igual que les variables meteorològiques, l'associació entre l'exposició a la contaminació atmosfèrica i l'agreuament de la malaltia, encara que en aquest cas no s'ha trobat una relació amb les morts, pot ser només una conseqüència de l'associació amb la incidència.

L'estudi mediambiental s'ha desenvolupat dins del subprograma 'Cohort-Real World Data' del CIBER de Epidemio-

logía y Salud Pública (CIBERESP) i ha rebut finançament parcial i incondicional per part del Fondo SUPERA COVID-19 (SAUN: Santander Universidades-CRUE-CSIC) i per part de COVID-19 Competitive Grant Program (Pfizer Global Medical Grants). Tanmateix va rebre finançament, en forma de cessió gratuïta de les dades, per part de l'AEMET.

Per a més detalls, consulteu l'Annex III.

Model de Mobilitat

El model epidemiològic fa ús les dades de mobilitat per caracteritzar la dinàmica epidemiològica, concretament les dades de mobilitat tenen impacte en les transicions entre susceptibles (S) i exposats (E), juntament amb altres paràmetres. Partint de les matrius de viatges que el MITMA posa a disposició es fa necessari calcular els paràmetres de mobilitat que es fan servir per modelar la mobilitat i que s'integren dins el model.

El model utilitza principalment dos grups de paràmetres, per una banda els paràmetres que ens indiquen quina proporció de la població que resideix en una determinada àrea no es mou de l'àrea (q), quina proporció de la població es mou dins l'àrea (r) i quina proporció es mou a altres àrees (p). Per altra banda també calculem la matriu de fluxos de mobilitat entre àrees, això és la probabilitat que donada una àrea una persona es mogui a una altra àrea (M_{ij}).

En el model epidemiològic, per tal de poder mantenir la coherència amb les dades de salut, només s'utilitza els viatges que es fan des del lloc de residència. Considerem només els trajectes que es fan amb destí la feina, d'aquesta forma intentem aproximar-nos a la mobilitat més important, la mobilitat recurrent. El lloc de residència i de feina, venen directament determinats per les dades que el MITMA ens proporciona i està degudament documentat en la metodologia referenciada. Les dades de mobilitat, que en les col·leccions en brut tenen una freqüència horària, són agregades per dia, de forma que tenim els valors dels paràmetres i una matriu de fluxos per cada dia.

Les àrees en les quals, en principi, treballem són les que ens proporciona el MITMA, àrees de mobilitat que venen determinades per municipis o districtes. Per tal de treballar a la mateixa escala espacial que el model epidemiològic fem una superposició de les capes que distribueix de forma lineal els viatges que es realitzen entre capes MITMA a les capes ABS amb les que treballa el model.

Model Epidemiològic Mecànic

El model epidemiològic que proposem està basat en un model compartimental. En particular, aquest model consta de vuit compartiments (veure Figura 13) que permeten estudiar les transicions entre les sub-poblacions (per a cada ABS) de persones susceptibles (S), exposades (E), infectades asimptomàtics (A), infectades simptomàtics (I), hospitalitzades (H), hospitalitzades en unitats de cures intensives (U), recuperades (R), i finades (D). Els compartiments d'infectats asimptomàtics i

simptomàtics junts formen el comunament denominat compartiment d'infectats. Es proposa aquesta segmentació per tal de caracteritzar amb precisió la dinàmica de la COVID-19.

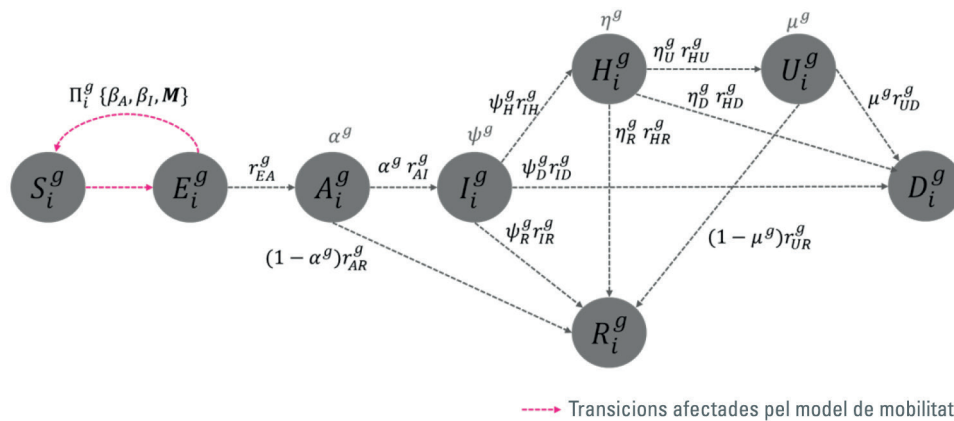


Figura 13. Model compartimental

Aquest model té tantes capes com sub-poblacions. Aquestes sub-poblacions s'han definit per tal de representar de la forma més adient possible la dinàmica d'infecció de la COVID-19. Tal com es mostra a l'últim informe sobre la situació de la COVID-19 a Espanya², l'afectació de la COVID-19 varia de forma significativa en funció del grup d'edat. Conseqüentment, les sub-poblacions venen donades per les dades de mobilitat i pels grups d'edat. D'aquesta manera podrem agrupar tota la població catalana demogràficament segons l'ABS a la qual pertanyen i segons el seu grup d'edat.

Anàlogament, la transició entre els diferents compartiments també dependrà del grup d'edat. Per exemple, basant-nos en la informació publicada fins al moment, la transició del compartiment de persones hospitalitzades (H) al compartiment de persones ingressades en la unitat de cures intensives (U) serà més probable pels grups de més edat que per la resta de la població.

Com es pot veure a la representació gràfica del model, un cop una persona està exposada a la COVID-19, aquesta o bé acabarà recuperant-se (R) o finalada (D). Alhora, pot haver persones que encara no hagin estat exposades a la COVID-19 i romanguin susceptibles (S).

Al gràfic es pot veure com cada transició entre compartiments té associat un paràmetre. Aquests paràmetres tenen el següent significat:

- **Transició S-E.** π_i^g descriu la força d'infecció, que és la probabilitat amb la qual una persona susceptible (S) del grup d'edat g de l'ABS i passarà al compartiment d'exposats (E) degut a un contacte amb una persona infecciosa (persones infectades asimptomàtiques (A) i simptomàtiques (I)).
- **Transició E-A.** r_{EA}^g descriu el període latent, és a dir, el temps que una persona triga en passar d'estar exposada (E) a ser considerada infecciosa asimptomàtica (A).
- **Transició A-I.** α^g representa la probabilitat de què les persones asimptomàtiques (A) del grup d'edat g desenvolupin símptomes (I), i r_{AI}^g descriu el temps que una persona triga en transicionar entre els dos compartiments esmentats.
- **Transició A-R.** $1 - \alpha^g$ representa la probabilitat de què les persones asimptomàtiques (A) es recuperin (R), i r_{AR}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.
- **Transició I-H.** ψ_H^g representa la probabilitat de què les persones infectades amb símptomes (I) siguin hospitalitzades (H).

2 Informe anomenat "Situación de COVID-19 en España a 13 de enero de 2021. Equipo COVID-19. RENAVE. CNE. CNM (ISCIII)" que es pot trobar al [següent enllaç](#) que pertany a la web de l'ISCIII.

zades (H), i r_{IH}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.

- **Transició I-D.** ψ_D^g representa la probabilitat de què les persones infectades amb símptomes (I) morin (D), i r_{ID}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.
- **Transició I-R.** ψ_R^g representa la probabilitat de què les persones infectades amb símptomes (I) es recuperin (R) sense necessitat de ser hospitalitzades, i r_{IR}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.
- **Transició H-U.** η_U^g representa la probabilitat de què les persones hospitalitzades (H), i r_{HU}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.
- **Transició H-D.** η_D^g representa la probabilitat de què les persones hospitalitzades (H) morin (D), i r_{HD}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.
- **Transició H-R.** η_R^g representa la probabilitat de què les persones hospitalitzades (H) es recuperin (R) sense necessitat d'ingressar a la UCI, i r_{HR}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.
- **Transició U-D.** μ^g representa la probabilitat de què les persones ingressades a la UCI (U) morin (D), i r_{UD}^g descriu

el temps que una persona triga en fer la transició entre els dos compartiments esmentats.

- **Transició U-R.** $1-\mu^g$ representa la probabilitat de què les persones ingressades a la UCI (U) es recuperin (R), i r_{UR}^g descriu el temps que una persona triga en fer la transició entre els dos compartiments esmentats.

La probabilitat d'infecció és un dels conceptes claus del model. La probabilitat d'infecció varia segons ABS i grup d'edat a cada moment. Això és degut a que el càlcul de la probabilitat d'infecció depèn, entre d'altres, de: la mobilitat, el nombre de contactes, el nombre de persones infeccioses que estiguin a l'ABS, el nombre de persones susceptibles que estiguin a l'ABS, i la capacitat d'infecció de les persones infectades, tant asimptomàtiques com simptomàtiques.

El nombre de contactes d'una persona diferirà segons el grup d'edat al qual pertanyi. Així, el càlcul del nombre total de contactes d'una persona s'obté inferint les interaccions entre els diferents grups d'edat.

Per tal d'assegurar que el model és sòlid i que, en conseqüència, els resultats obtinguts són vàlids, s'ha analitzat el seu comportament de forma teòrica arribant a la fonamentació matemàtica del model. Per a més informació, consulteu l'Annex III.

Model Epidemiològic Dirigit per Dades

L'inici de les activitats dirigides per dades de la iniciativa estaven previstes per la segona fase del projecte però atenent a que durant la primera fase del projecte ja ha quedat disponible el servei COVID-19 Flow Maps, hem pogut emprar les dades d'aquest servei a partir de la seva API de connexió i integrar les dades diàries d'infecció i la mobilitat entre àrees. En aquesta primera aproximació a

la realització de prediccions de l'evolució de la pandèmia hem treballat a nivell nacional Espanyol. Per tant, hem agregat el risc associat a la mobilitat per província a fi i efecte d'avaluar el risc de propagació a nivell nacional. Aquest estudi s'ha realitzat amb una xarxa del recurrent del tipus LSTM de tres capes.



Anàlisi



Estudi de la mobilitat i els determinants socioeconòmics

L'estudi de la mobilitat és fonamental per entendre com evoluciona un procés epidèmic ja que pot ser un indicador important dels contactes entre individus i, per tant, de la potencial transmissió del virus. En endavant fem una anàlisi descriptiva de la mobilitat a Catalunya des del dia 14 de febrer fins al dia 31 de desembre de 2020. Aquesta anàlisi abasta les diferents etapes de l'epidèmia (primera onada, segona onada, període de transició entre onades) amb totes les seves respectives fases (restriccions de mobilitat, estats d'alerta, etc.).

Els paràmetres de mobilitat amb els que treballarem per analitzar diferents situacions són: (1) índex de mobilitat:

és la proporció de persones que fan un o més viatges durant un determinat dia, (2) increment de la mobilitat: és la diferència en el nombre de persones que fan un o més viatges durant un dia en relació amb la referència, les dues setmanes que van des del 17 de febrer fins al 1 de març, i (3) el nombre de viatges totals que s'han realitzat durant un determinat dia.

A més de l'estudi de la mobilitat, presentem una anàlisi de les relacions que existeixen entre la mobilitat i algunes variables socioeconòmiques, principalment la renda, en el context de l'epidèmia i en cada una de les seves fases.

	Febrer	Març	Abril	Maig	Juny	Juliol	Agost	Setembre	Octubre	Novembre	Desembre
Restriccions de mobilitat		15	25								
Mobilitat essencial		50	7								
Restriccions de mobilitat: fase 0			10	10							
Desescalada				15							
Mobilitat sense restriccions					19						
Situació de partida en alerta 5									25		
Tram 1 de l'alerta 5										23	
Tram 1 de l'alerta 5 amb les mesures de flexibilització											14
Mesures de contenció Nadal, Cap d'Any i Reis											23

Figura 14. Calendari de mesures que afecten la mobilitat

Evolució de la mobilitat

La mobilitat a Catalunya ha anat variant de manera molt significativa des de l'inici de l'epidèmia a causa de l'aplicació de mesures de salut pública no farmacològiques que han impactat directament en la mobilitat. En general podem veure com l'índex de mobilitat diari previ a la pandèmia se situa al voltant de 0.7 i segueix els patrons periòdics setmanals habituals. Durant els dies feiners l'índex de mobilitat es manté a nivells alts i pràcticament constants, representant la mobilitat que es produeix pels desplaçaments laborals i altres activitats recurrents. La mobilitat disminueix els dissabtes i, de forma més significativa, els diumenges o festius.

A mesura que es van aplicant restriccions a la mobilitat

aquest índex s'ha vist reduït, a vegades de manera molt significativa, això es fa evident especialment durant la primera onada, que analitzarem en més detall, però també a partir del mes d'octubre, especialment els caps de setmana. En general, però, l'índex de mobilitat durant els dies feiners no ha variat de forma dràstica des de la setmana del 22 de juny, quan es varen acabar les restriccions de mobilitat. Tot i això, encara que la proporció de la població que es mou sigui similar durant aquesta etapa, des de 22 de juny fins al 31 de desembre, els patrons de mobilitat i els contactes que es produeixen són diferents. Els contactes varien en funció de l'entorn i l'activitat social, per exemple, durant l'estiu la mobilitat i els contactes són diferents que durant el setembre quan comencen les classes a les escoles o les universitats, el mateix suc-

ceix quan la gent s'incorpora a la feina després de les vacances d'estiu o d'un pont. Aquests contactes causats per l'activitat social no es veuen representats en aquesta

anàlisi, ja que són necessàries dades a un nivell molt més desagregat del que disposem.

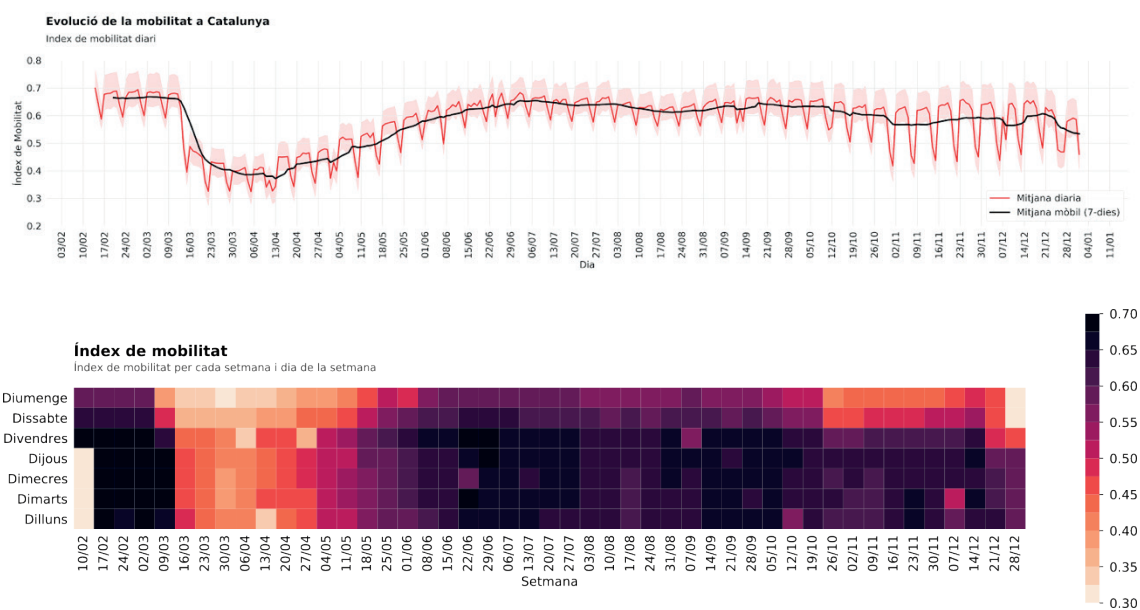


Figura 15. Evolució de l'índex de mobilitat des del 14 de febrer fins al 31 de desembre per tot Catalunya

Reducció de la mobilitat

Si analitzem la reducció de la mobilitat, en relació amb la mobilitat de referència, podem veure de manera més evident els canvis i l'impacte de les restriccions. En general s'observa com la mobilitat va caure dràsticament durant la primera onada en diferents nivells depenent de les mesures que s'estaven aplicant en cada moment. També apareixen canvis importants d'engà del 25 d'octubre quan la mobilitat

de cap de setmanes i divendres cau. A la Figura 15 també podem observar algunes caigudes de mobilitat puntuals, aquestes corresponen en general a dies festius entre setmana. Durant la primera onada destaquen el 10 (divendres sant) i 13 d'abril (dilluns de pasqua), 1 de maig i 24 de juny. Durant la segona onada l'11 de setembre, el 12 d'octubre, el pont de la immaculada i les festes nadalenques.

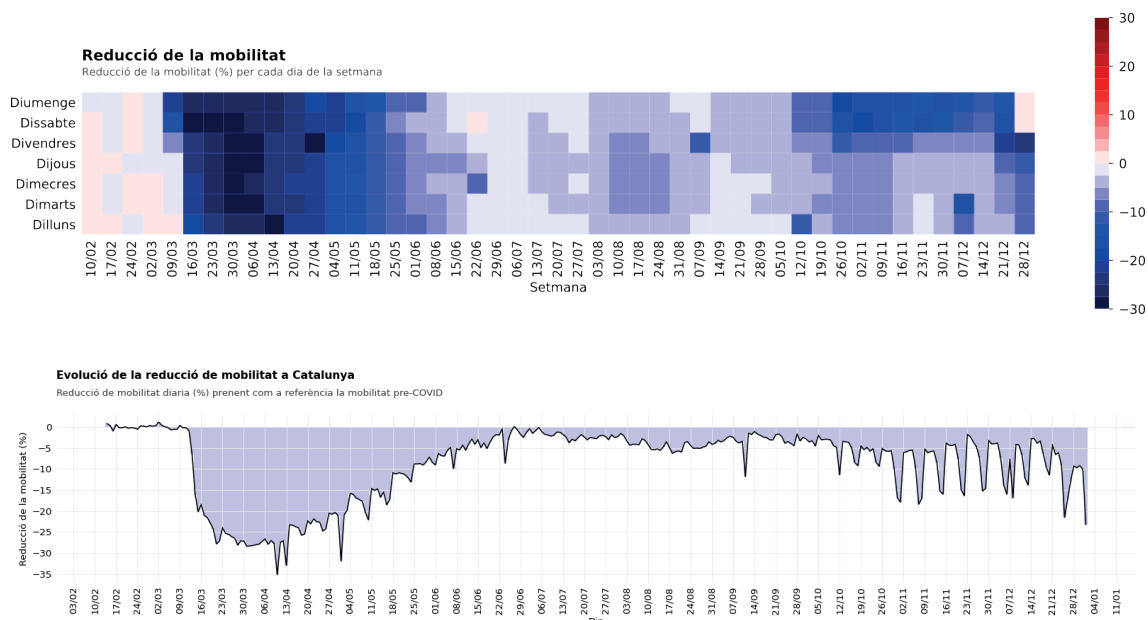


Figura 16. Evolució de la reducció de la mobilitat des del 14 de febrer fins al 31 de desembre per tot Catalunya

Si utilitzem com a paràmetre el nombre de viatges que es fan a Catalunya i els dividim segons l'origen i destí dels trajectes tenim: mobilitat interna, trajectes amb origen i destí Catalunya; mobilitat entrant, trajectes que tenen com a destí Catalunya amb origen fora de Catalunya; i mobilitat sortint, trajectes que tenen com a origen Catalunya i destí fora de Catalunya.

S'observa que la dinàmica que descriuen els viatges que es produeixen dins Catalunya és anàloga a què es descriu amb l'índex de mobilitat. Durant l'epidèmia els pics de mobilitat es produeixen a principis de juliol i a principis de setembre. A mitjans d'octubre el nombre de viatges cau lleugerament i torna a pujar a finals de novembre. En general, tot i que existeix una disminució de la mobilitat durant

el mes de novembre aquesta és molt petita si la comparem amb les reduccions de mobilitat de la primera etapa de l'epidèmia provocades per les mesures de salut pública.

Per altra banda també podem veure com les restriccions de mobilitat han afectat als viatges que es fan amb origen o destí Catalunya. Principalment el que es detecta de manera molt precisa és l'obertura que es va produir el mes de juny i el tancament que es va produir al mes d'octubre. La mobilitat d'entrada i sortida de Catalunya des de finals d'octubre se situa a nivells similars a la primera etapa de l'epidèmia durant els caps de setmana i es veu incrementada durant els dies laborables.

Una altra informació significativa que ens dona aquesta mètrica és la dinàmica de desplaçaments d'entrada i sortida durant períodes de vacances. Veiem molt clarament com durant les dues primeres setmanes d'agost es produeixen molts més viatges de sortida que d'entrada de Ca-

talunya i per altra banda com la segona quinzena d'agost es produeixen més viatges d'entrada. La mateixa dinàmica es repeteix per l'11 de setembre i el 9 d'octubre, on hi ha primer molts viatges de sortida i posteriorment d'entrada a Catalunya.

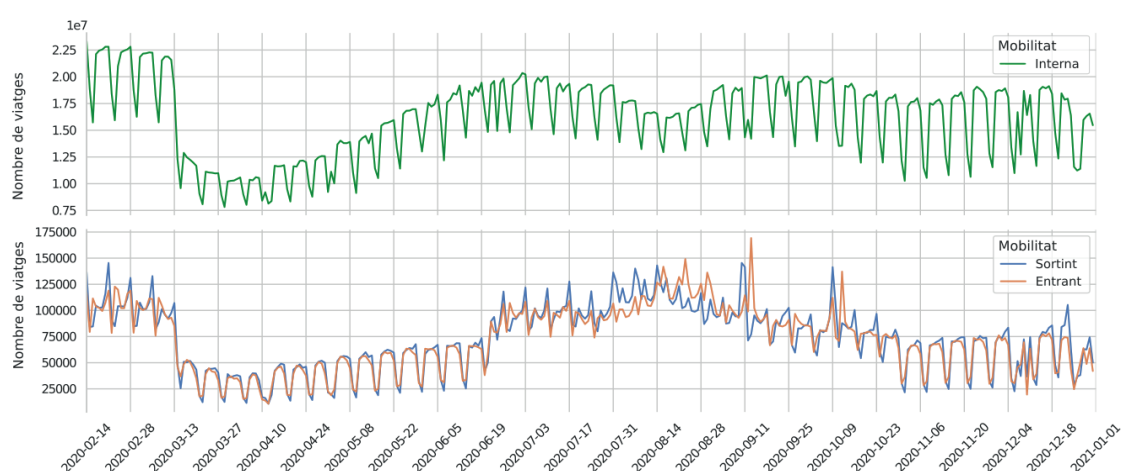


Figura 17. Evolució del nombre de viatges dins Catalunya (mobilitat interna) i d'entrada i sortida a Catalunya des del 14 de febrer fins al 31 de desembre.

Mobilitat durant la primera etapa

Ens centrem en la primera etapa de l'epidèmia, podem dividir les mesures que impacten en la mobilitat en quatre fases: (1) s'apliquen restriccions de mobilitat, està permesa només la mobilitat laboral i essencial; (2) només està permesa la mobilitat essencial; (3) està permesa la mobilitat laboral i essencial i (4) desescalada obertura de la mobilitat que es va fent de forma esglaonada a diferent ritme en funció de la incidència a cada àrea sanitària.

Si agafem la mobilitat de les dues setmanes que van des del 17 de febrer fins a l'1 de març com a referència podem veure l'impacte a l'índex de mobilitat per cada dia de la setmana i el coeficient de correlació entre la mobilitat de referència i la del període corresponent. A la Figura 18 podem veure com a mesura que s'han aplicat les restriccions la mobilitat s'ha vist reduïda i a mesura que s'ha dut a terme la desescalada s'ha anat recuperant.

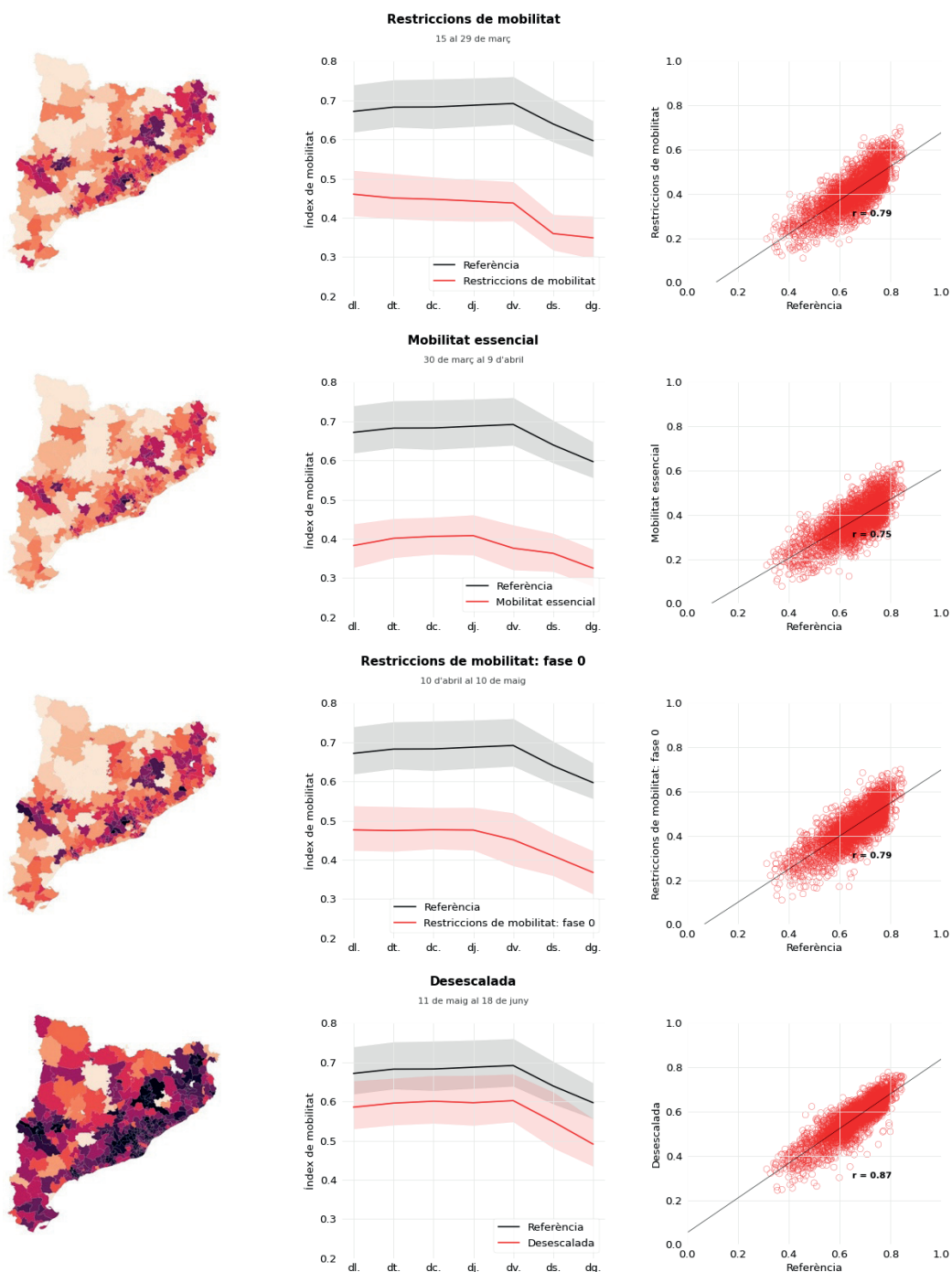


Figura 18. Índex de mobilitat durant les diferents etapes d'aplicació de restriccions de mobilitat a la primera onada. El mapa de la columna esquerra representa l'índex de mobilitat d'un dilluns (23 de març, 6 d'abril, 20 d'abril i 25 de maig) dins el període de restriccions, la figura central és l'índex de mobilitat mig per cada dia de la setmana i el gràfic de la columna dreta mostra la dispersió àrea a àrea entre la mobilitat de referència i la de cada etapa, juntament en el coeficient de correlació.

Mobilitat i factors socioeconòmics

Les dinàmiques de mobilitat individual vénen determinades per factors molt diversos, especialment durant una epidèmia on les polítiques de salut pública incideixen directament en ella. Considerant que el confinament és una de les mesures excepcionals que afecta la reducció de la mobilitat i que la mobilitat està molt lligada a l'activitat laboral, sembla que els factors socioeconòmics (com la renda, l'atur, el tipus d'activitat econòmica, tipus

d'habitatge, etc.) poden tenir un cert pes.

En aquest cas ens centrarem a estudiar les correlacions que existeixen entre la renda mitjana per càpita i índex de mobilitat en una determinada àrea (municipi o districte) per les diferents fases de la primera etapa de l'epidèmia. Aquest anàlisi el durem a terme a escala provincial.

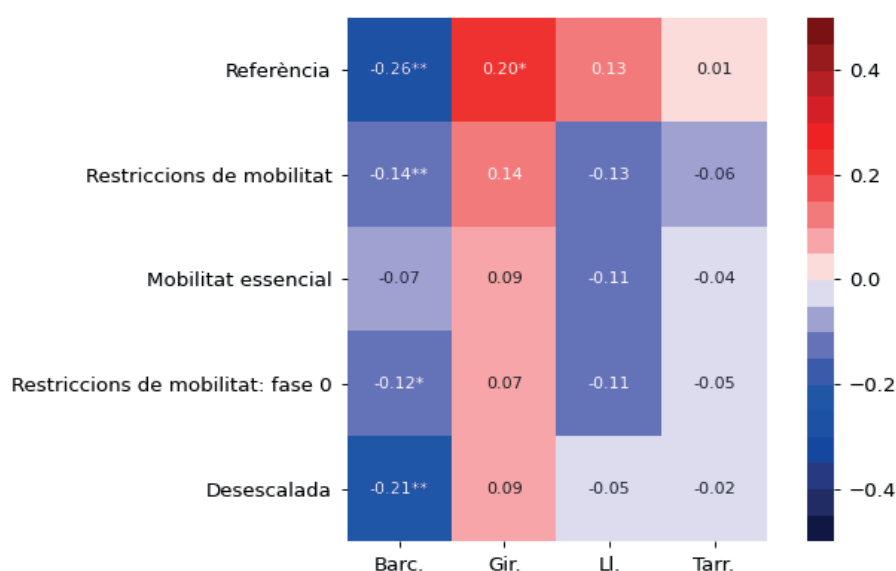


Figura 19. Correlació entre l'índex de mobilitat i la renda mitjana per persona a Barcelona, Girona, Lleida i Tarragona durant les diferents fases de la primera onada. El valor indica el coeficient de correlació i la significança ve donada per ** ($p < 0.05$) i * ($p < 0.1$).

La correlació entre mobilitat i renda varia en funció de l'entorn on aquesta es produeix; es pot observar que durant el període de referència, que considerem com situació de normalitat, ja que no hi ha cap mesura que afecti la mobilitat, la mobilitat entre Barcelona i Girona (on la correlació és significativa), la mobilitat correlaciona de manera inversa. És a dir, a Barcelona, com major és la renda menor és la mobilitat i a Girona, com major és la renda major és l'índex de mobilitat.

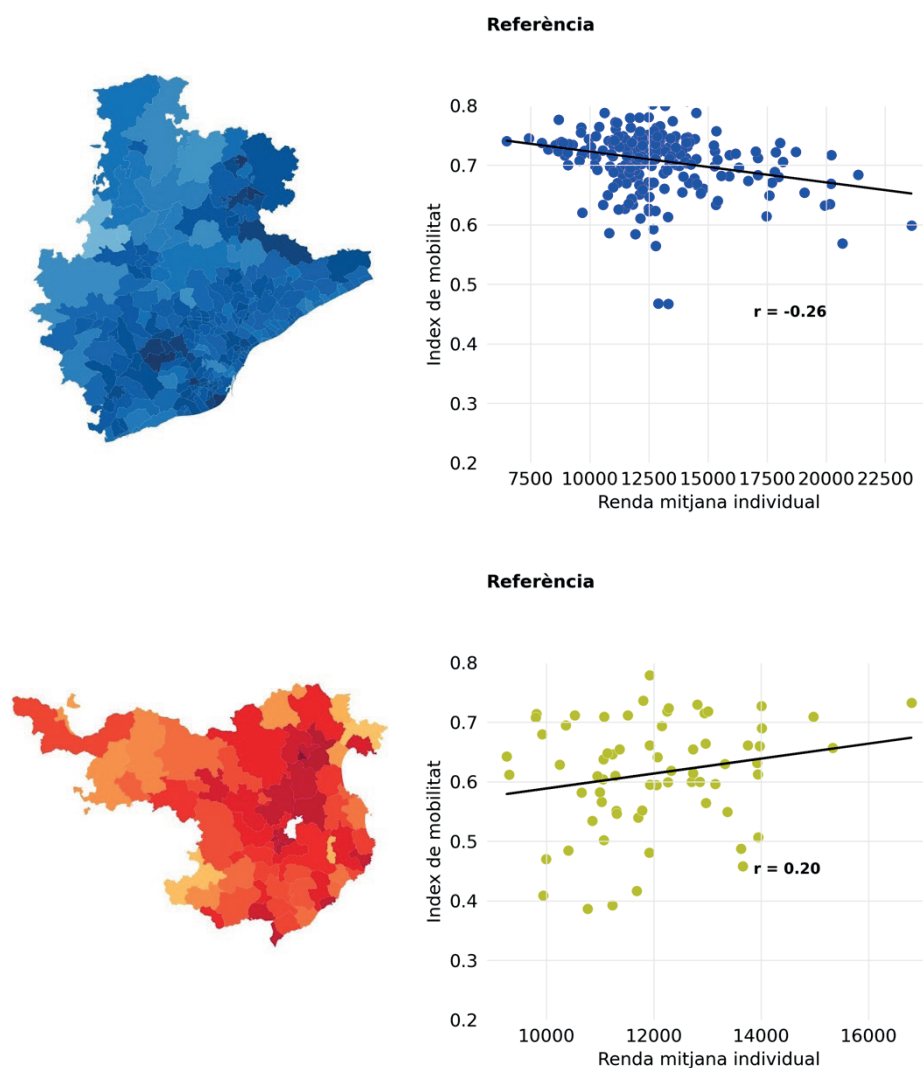


Figura 20. Correlació entre l'índex de mobilitat i la renda mitjana per persona a Barcelona i Girona des del 17 de febrer a l'1 de març. El mapa de la dreta representa la mobilitat mitjana durant aquell període.

En aplicació de mesures restrictives de mobilitat només tenim resultats significatius a la província de Barcelona durant les restriccions de mobilitat i la desescalada, però no durant les restriccions de mobilitat essencial. Podem veure que a mesura que s'apliquen mesures l'índex de correlació va disminuint de manera que les diferències entre la mobilitat de les persones amb més renda i en menys és menor, tot i això la tendència és la mateixa que en situació de normalitat.

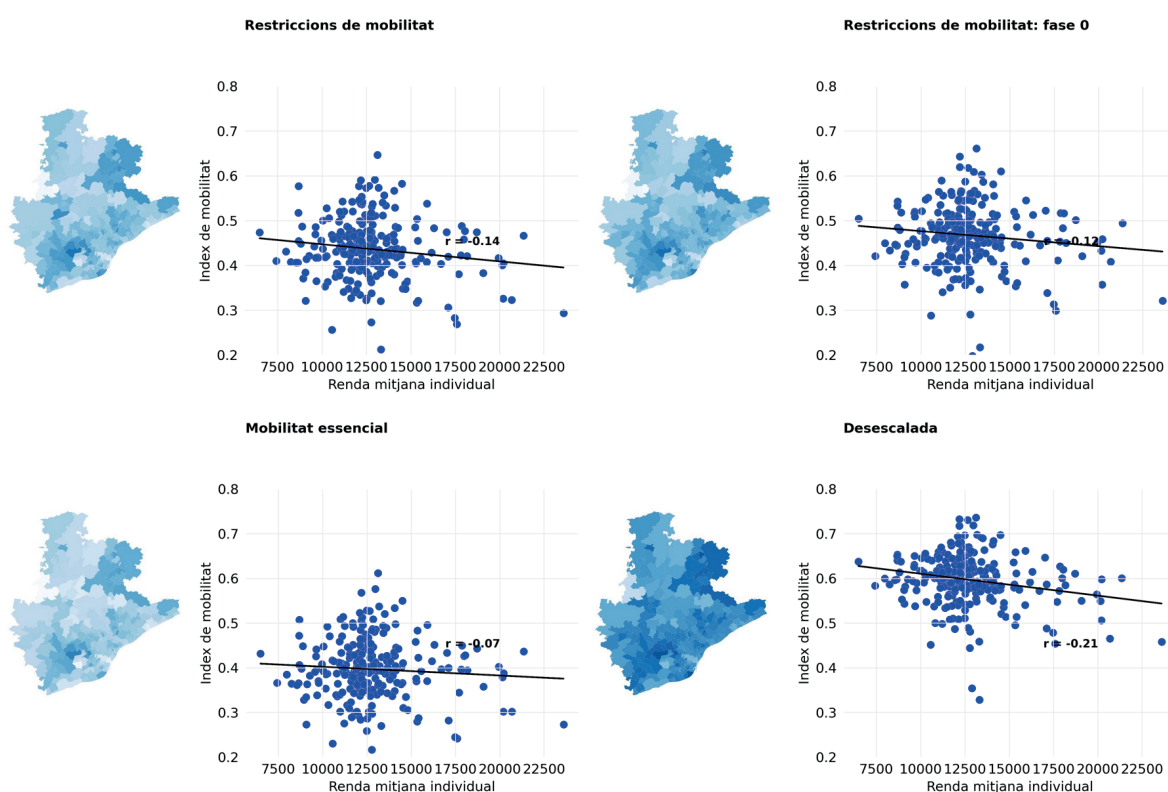


Figura 21. Correlació entre l'índex de mobilitat i la renda mitjana per persona a Barcelona durant la primera onada. El mapa de la columna esquerra representa l'índex de mobilitat mitjana dins el període corresponent i el gràfic de la columna dreta mostra la dispersió entre índex de mobilitat i la renda mitjana per persona a cada àrea, juntament amb el coeficient de correlació.

Mobilitat durant la segona etapa

Des del 22 de juny fins als 25 d'octubre no es van aplicar mesures que afecten la mobilitat general de Catalunya. A nivell agregat l'índex de mobilitat s'ha mantingut a nivells alts amb petites pujades i baixades durant l'estiu i durant els mesos de setembre i octubre. Tot i això s'ha de tenir en compte que els contactes durant els mesos de juliol i agost són diferents que el setembre i octubre, l'activitat lectiva a les escoles, instituts i universitat, i l'activitat laboral fan que les dinàmiques d'interaccions socials siguin diferents, ja que es donen en entorns i entre diferents persones. Amb les dades de mobilitat agregades que disposem no és possible arribar al detall per aclarir quins són aquests contactes.

Les mesures que impacten en la mobilitat durant la segona etapa comencen a aplicar-se el dia 25 d'octubre. Aquestes mesures afecten principalment a la mobilitat durant la nit (toc de queda) i els caps de setmanes (tancament municipal). L'índex de mobilitat, que té en compte la població que com a mínim fa un viatge al dia, cau durant els caps de setmanes, on es pot apreciar una baixada considerable de la mobilitat respecte de la mobilitat de referència.

Si comparem la mobilitat en estat d'alerta a Catalunya en la fase 0 i la fase 1 veiem que a escala general pràcticament no s'ha vist modificat, tot i això no podem dir que els contactes que es produeixen siguin de la mateixa intensitat.

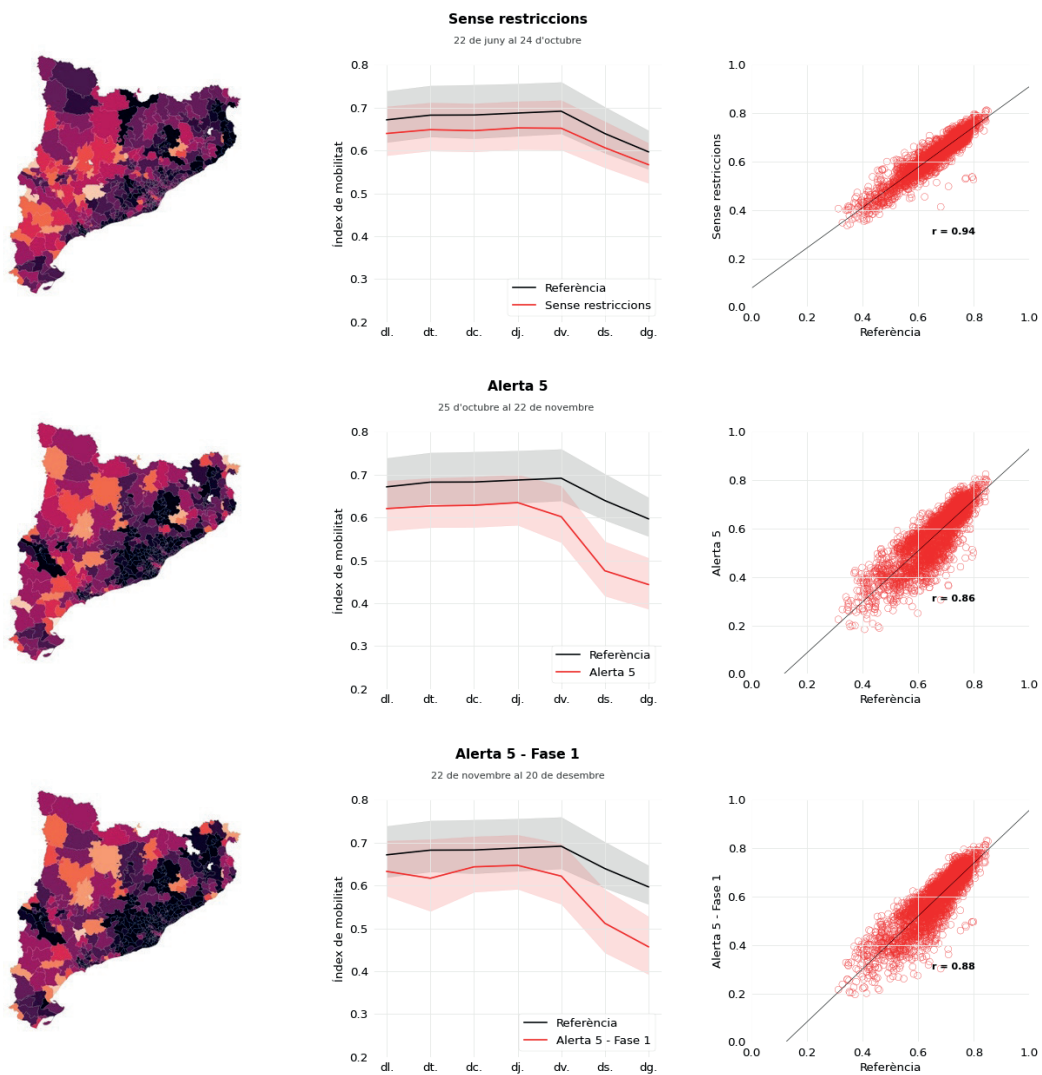


Figura 22. Índex de mobilitat durant la segona fase, inclòs el període sense restriccions. El mapa de la columna esquerra representa l'índex de mobilitat d'un dilluns (5 de juliol, 9 de novembre i 30 de novembre) dins el període de restriccions, la figura central és l'índex de mobilitat mig per cada dia de la setmana i el gràfic de la columna dreta mostra la dispersió àrea a àrea entre la mobilitat de referència i la de cada etapa, juntament amb el coeficient de correlació.

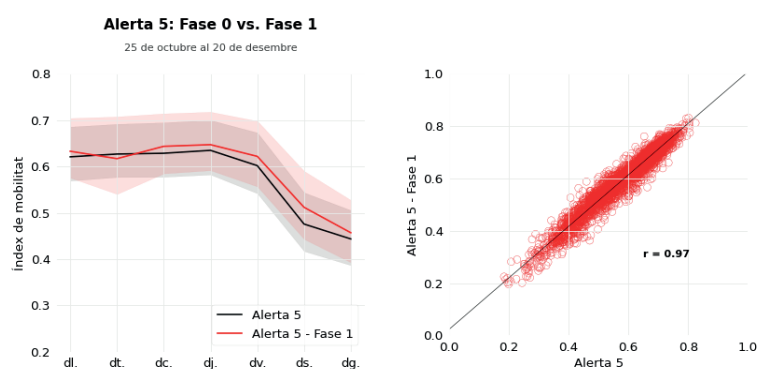


Figura 23. Diferències entre l'índex de mobilitat les restriccions en Alerta 5.

Mobilitat i factors socioeconòmics

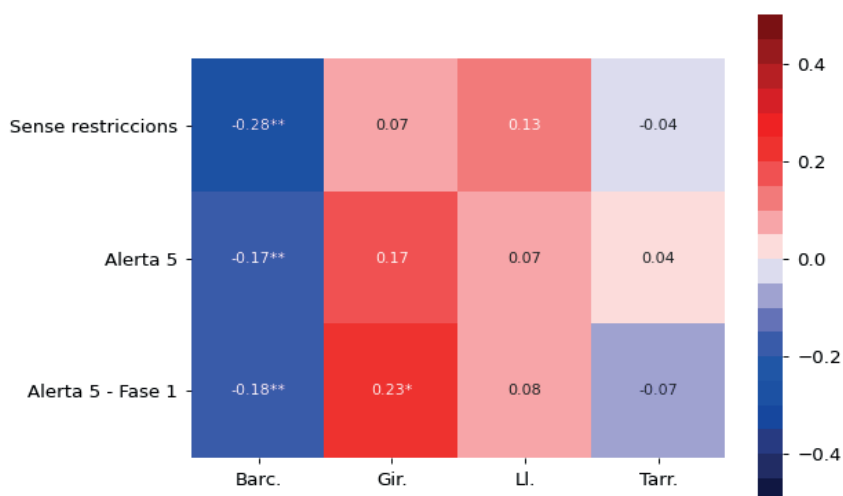


Figura 24. Correlació entre l'índex de mobilitat i la renda mitjana per persona a Barcelona, Girona, Lleida i Tarragona durant les diferents fases de la segona onada. El valor indica el coeficient de correlació i la significança ve donada per ** ($p < 0.05$) i * ($p < 0.1$).

En aquesta segona etapa l'índex de mobilitat no presenta grans variacions entre les diferents fases i, per tant, la correlació entre la mobilitat i la renda és similar entre la fase 0 i 1 de l'alerta 5 a Barcelona.

De la mateixa manera que en l'anàlisi de la primera etapa, que en situació d'alta mobilitat (sense restriccions) la correlació és significativa això succeeix en les tres etapes.

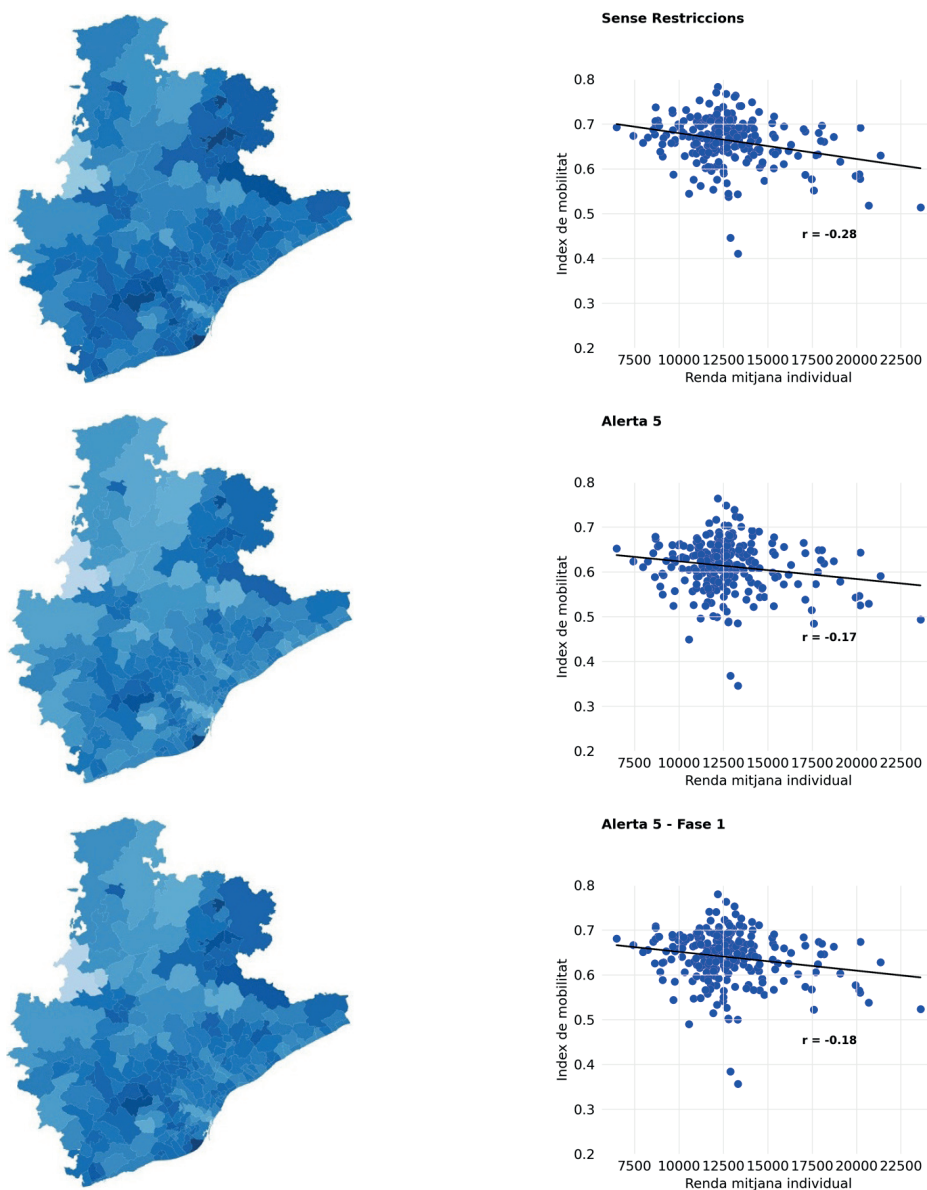


Figura 25. Correlació entre l'índex de mobilitat i la renda mitjana per persona a Barcelona durant la segona onada. El mapa de la columna esquerra representa l'índex de mobilitat mitjana dins el període corresponent i el gràfic de la columna dreta mostra la dispersió entre índex de mobilitat i la renda mitjana per persona a cada àrea, juntament amb el coeficient de correlació.

A la província de Girona la dinàmica general és just la inversa, com més alta és la renda mitjana individual d'una zona, més alta és la mobilitat, i això és lleugerament significatiu ($p < 0.1$) en el cas de l'alerta 5 (fase 1). Aquesta diferència relativa entre alerta 5 i alerta 5 (fase 1) pot ser

causada per diferents motius, pot ser una conseqüència directa de les mesures de salut pública, però també pot ser pel calendari festiu i de ponts que cau dins aquesta fase i provoca un augment de viatges a determinades zones, particularment les que tenen un alt poder adquisitiu.

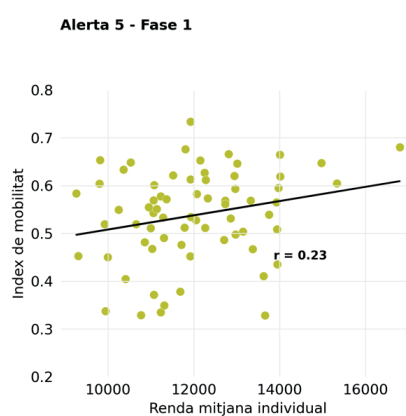
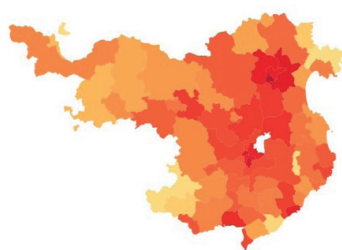


Figura 26. Correlació entre l'índex de mobilitat i la renda mitjana per persona a Girona durant la segona onada. El mapa de la columna esquerra representa l'índex de mobilitat mitjana del període corresponent (Alerta 5 – Fase 1) i el gràfic de la columna dreta mostra la dispersió entre índex de mobilitat mitjana i la renda mitjana per persona a cada àrea, juntament amb el coeficient de correlació.

Anàlisi dels desplaçaments durant el pont de desembre

Les dades de mobilitat ens permeten estudiar esdeveniments que generen una mobilitat especial i que durant el procés epidèmic són molt rellevants. Podem veure que durant el mes de desembre hi ha diferents moments en què es produeixen augments i disminucions de la mobilitat

anòmals. En general són dies de festa o ponts on les dinàmiques de mobilitat són diferents, ja que l'activitat escolar, acadèmica i laboral es veu modificada. En aquest cas centrem l'anàlisi en el pont del 6 i 8 de desembre.

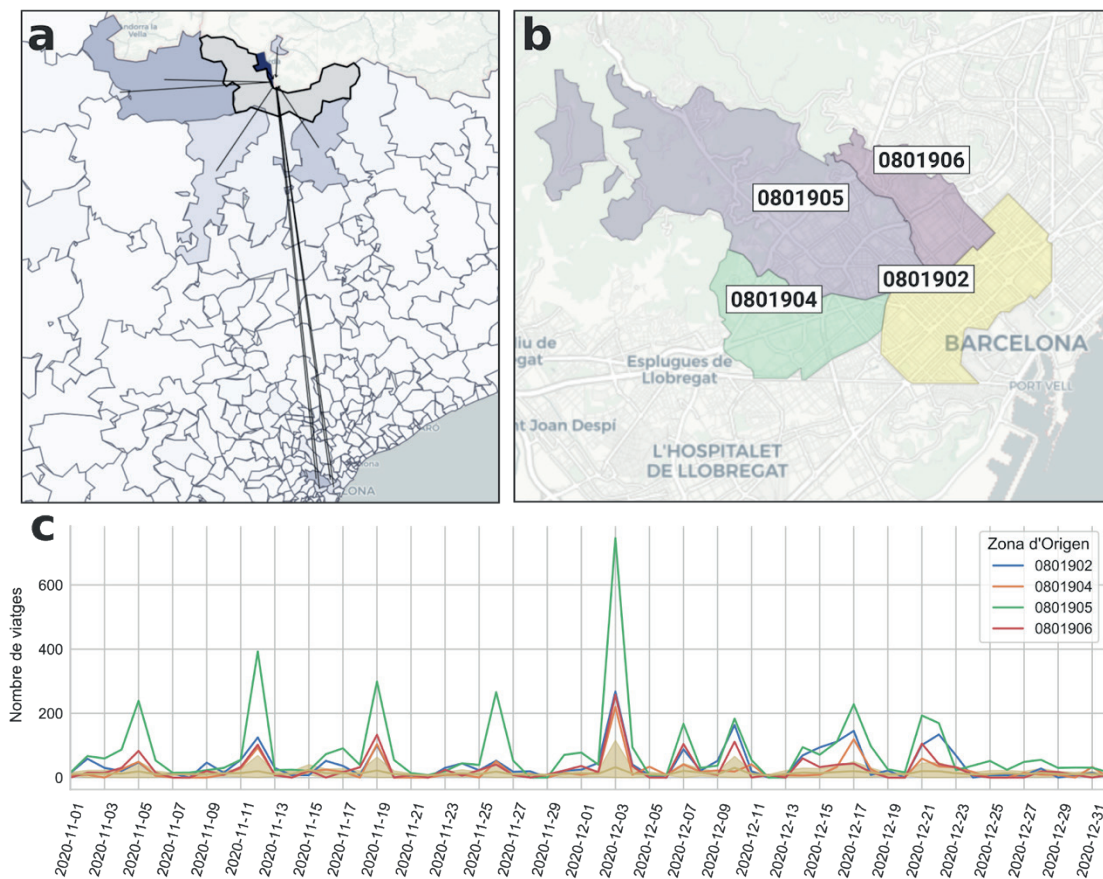


Figura 27. Moviments anòmals durant el pont de desembre. La figura (a) representa les àrees amb major mobilitat a la Cerdanya el dijous 3/12 previ al pont. En la figura (b) es detallen quines són les zones de Barcelona que corresponen als orígens dels viatges destacats com anòmals: Sarrià-Sant Gervasi (0801905), les Corts (0801904), Gràcia (0801906) i l'Eixample (0801902). A la figura (c) es representa la sèrie temporal del nombre de viatges amb destí la Cerdanya considerant com a origen les zones que estan a més de 50 km i que mostren un major nombre de viatges que la mitja i una desviació estàndard. Les distintes corbes corresponen a les sèries temporals del nombre de viatges amb origen a les zones destacades com anòmals i l'àrea groga correspon a la mitjana.

Durant el pont es va detectar un nombre molt elevat de viatges al Pirineu, més concretament a la Cerdanya. Si estudiem la distribució estadística dels viatges considerant i no considerem els viatges de les zones adjacents, que són les que tenen més intercanvis, en trobem que els orígens amb més destí als Pirineus els trobem a Barcelona. Concretament s'identifiquen quatre districtes que exhibeixen un nombre de viatges molt elevat i que s'allunya significativament de la mitjana, aquests quatre districtes són: Sarrià-Sant Gervasi, les Corts, Gràcia i l'Eixample. És notable

que una zona en particular, Sarrià-Sant Gervasi, va registrar un nombre superior a 700 viatges durant el dijous previ al pont significativament més elevat que les altres àrees. De fet, podem veure com el dijous previ al pont, durant el mes de novembre quan s'aplicaven restriccions municipals de cap de setmana, es van produir una quantitat important de viatges en aquesta zona del Pirineu des de Barcelona de manera periòdica. Cal recalcar que, aproximadament, 10 dies després del pont la zona analitzada del Pirineu va ser confinada degut a l'impacte de l'epidèmia.

Simulacions d'escenaris pandèmics amb el model epidemiològic mecànic

Per tal de facilitar l'entesa del model epidèmic plantejat, a continuació mostrem les gràfiques resultants d'executar el model sobre una població amb una única ABS i un únic grup d'edat formada per 101 persones durant la primera onada de la pandèmia. Inicialment, ens trobem amb 100 persones susceptibles (S) i 1 infectada que mostra símptomes (I). La següent gràfica mostra el resultat del model aplicat a aquesta població durant 100 dies. Inicialment,

la majoria de persones són susceptibles i, degut als contactes de la persona infecciosa amb la resta de la població, poc a poc, el nombre de persones recuperades (R) augmenta. Un cop passats els 100 dies veiem com la majoria de la població ha estat infectada i s'ha recuperat, però hi ha una petita part de la població que roman susceptible (S).

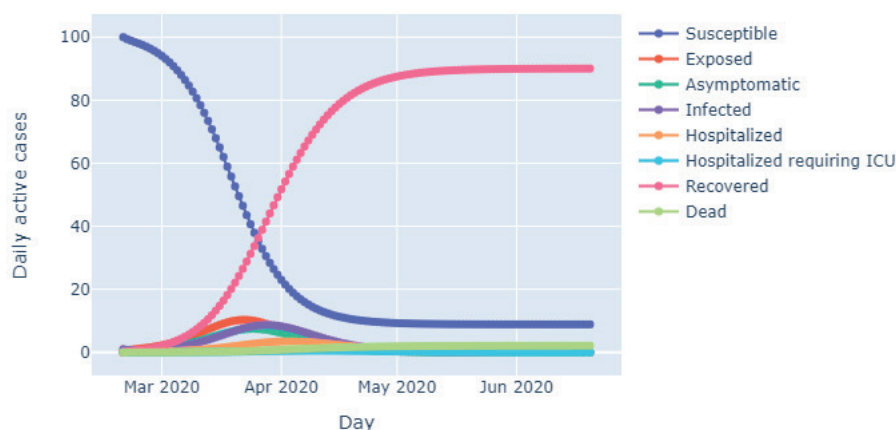


Figura 28. Resultat del model aplicat sobre una població amb una única ABS i un únic grup d'edat formada per 101 persones durant la primera onada de la pandèmia.

La següent gràfica mostra la quantitat de persones que cada dia entren a cada compartiment. Podem veure com el compartiment que creix més aviat és el de persones exposades (E), després el d'infectades asimptomàtiques (A), i a continuació el d'infectades simptomàtiques (I). Aquest comportament és l'esperat ja que la probabilitat de canviar entre aquests compartiments és de pràcticament 1.

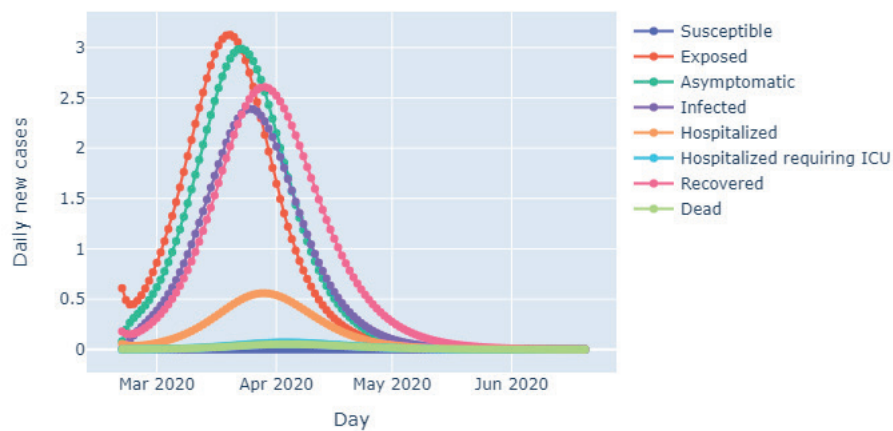


Figura 29. Resultat del model: nombre de persones que cada dia entren a cada compartiment

Per últim, la següent imatge mostra l'evolució dels cassos acumulats per a cada compartiment. Al final de la simulació, el nombre de persones susceptibles (S), recuperades (R), i finades (D) sumarà 101, que era la mida de la població d'estudi.

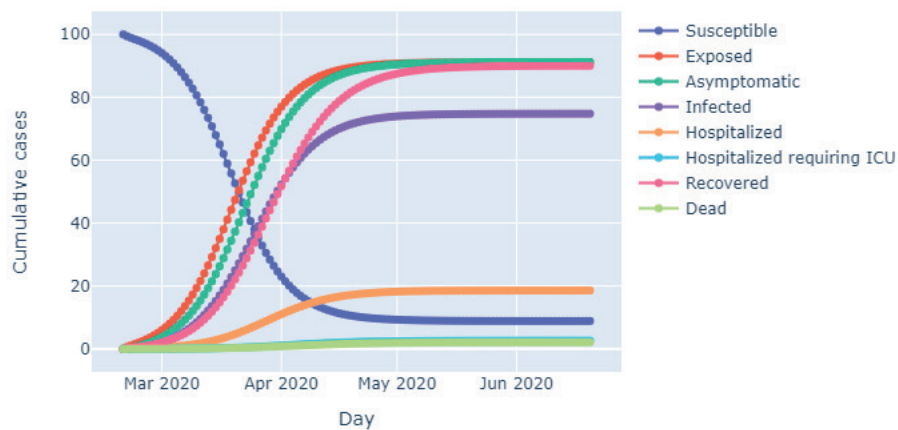


Figura 30. Resultat del model: l'evolució dels cassos acumulats per a cada compartiment.

Descripció i resultats del calibrat

El model epidèmic es fonamenta sobre diferents paràmetres que descriuen la dinàmica d'infecció i d'evolució de la COVID-19. Cadascuna de les onades que hem viscut fins al moment té unes peculiaritats diferents que la resta. Aquestes podríem agrupar-les segons el seu tipus:

- **Dinàmiques de contagi:** ús d'equips de protecció individual, nombre de contactes, aïllament, restriccions de mobilitat, entre d'altres.
- **Característiques clíniques:** ocupació hospitalària, ocupació de les UCIs, entre d'altres.
- **Comptabilització de casos:** nombre de PCR's realitzades, cribratges per barris, ús de tests d'antígens, entre d'altres.

Per exemple, al principi de la primera onada, que va començar al primer trimestre de 2020, encara no s'usava cap equip de protecció individual, poc després es va implantar el confinament domiciliari, i l'ocupació de les UCIs va arribar al seu màxim, impedit que gent que podria necessitar aquestes cures intensives pogués gaudir-les. A la segona onada es van detectar molts més casos positius degut, entra d'altres, a l'increment en la disponibilitat de proves PCR, i també van variar les dinàmiques de contagi en relació a la primera onada ja que, durant un temps, el nombre màxim de persones que es podien reunir va arribar a les 20 persones. I, per últim, la tercera onada es caracteritza pel toc de queda, les restriccions de mobilitat, i els cribratges per barris.

Per a poder usar el model per a fer prediccions a futur, cal que el model empli els paràmetres adequats per a descriure el que es vol representar. El procediment per a obtenir els paràmetres adients per a descriure una situació en concret s'anomena *calibrat*. Seguint amb la lògica exposada al paràgraf anterior, es calibrarà cada onada per separat. Així, cada onada es modelitzarà tenint en compte les seves particularitats i obtindrem resultats més acurats que si ho féssim d'altra manera.

Per a calibrar el model és necessari definir què és el que volem predir. En altres paraules, per a obtenir els paràmetres que millor representin la realitat que es vol modelitzar, cal definir com es mesura aquesta realitat. Podríem escollir entre: el nombre de persones infectades simptomàtiques (I), el nombre de persones hospitalitzades (H), el nombre de persones ingressades en UCIs (U), o bé el nombre de persones finades (D). Donada la baixa fiabilitat de moltes d'aquestes dades en les diferents onades, proposem calibrar el model usant el nombre de persones finades. Altra opció que pot semblar adient és l'ús dels ingressos en UCIs. Però, a la primera onada es va patir una saturació de les UCIs, fet que va portar als professionals sanitaris a haver de prioritzar l'ingrés de pacients d'urgència. Així, els registres d'ingressos en les UCIs de la primera onada no representen de forma fidedigna la severitat dels pacients d'aquell moment. En conseqüència, usarem el nombre de persones finades degut a la COVID-19. Considerem que les dades de mortalitat són fiables ja que aquestes són facilitades per funeràries.

Per a calibrar el model usem dades de l'afectació per COVID-19 a Espanya proveïdes pel *Ministerio de Sanidad, Consumo y Bienestar Social* del Govern espanyol³. Usem aquestes dades i no les dades d'afectació exclusives de Catalunya perquè ens interessa tenir una imatge global del què va passar a Espanya. Aquesta estratègia és la més adient degut a l'alta mobilitat entre Catalunya i Espanya. Si ens limitéssim a estudiar l'afectació a Catalunya, perdríem tot el context i la realitat catalana no es veuria degudament reflectida. Cal notar que es desconeix quin és el nombre exacte de persones asimptomàtiques a l'inici de la pandèmia. Analitzar les dinàmiques de la COVID-19 a Espanya ens permet arribar a aquest valor inicial de persones asimptomàtiques amb més facilitat que si només tinguéssim en compte el que passa a Catalunya. En altres paraules, aquesta mirada global ens permet entendre la realitat de la pandèmia a Catalunya.

3 **Plataforma del Ministerio de Sanidad, Consumo, y Bienestar Social** on es comparteixen les dades d'afectació per la COVID-19 a nivell espanyol.

Per a calibrar el model usem un algoritme basat en la *inferència bayesiana* que proposa els valors més adients per a cada paràmetre per tal d'ajustar-se al màxim a les dades a predir que hem definit anteriorment.

El calibrat d'aquest model és molt costosa a nivell computacional. Per exemple, la millor calibrat fins al moment resulta en un error mitjà de l'1.79% en el nombre de persones finades entre el 20 de febrer de 2020 i el 21 de

maig del 2020, que correspon a 500 morts, i va necessitar 36h de computació.

La següent gràfica mostra els resultats de dos calibrats de la primera onada: la primera mostrada en lila, i la segona en vermell, i les dades observades en color verd. Com es pot veure, el segon calibrat (vermell) és la que resulta en millors resultats ja que és la que millor s'ajusta a les dades reals.

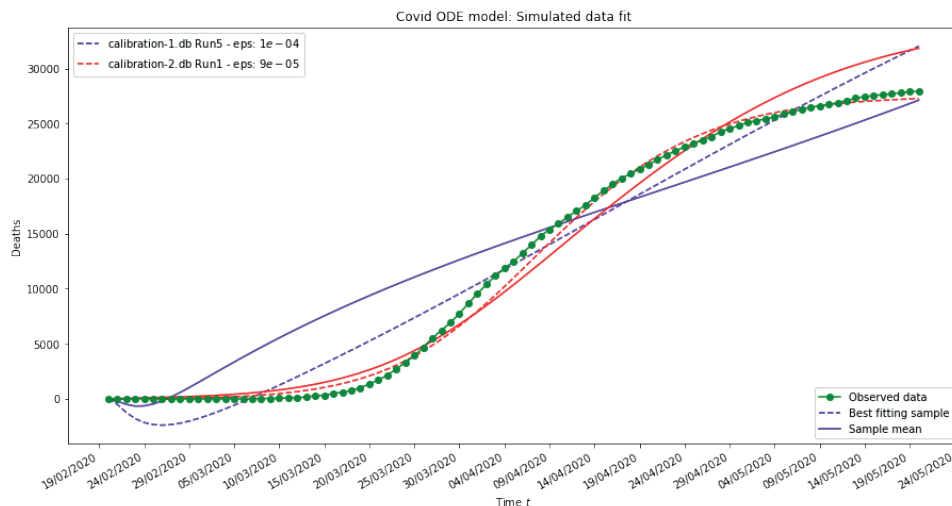


Figura 31. Resultats de dos calibrats de la primera onada: la primera mostrada en lila, i la segona en vermell, i les dades observades en color verd.

Els errors obtinguts en aquests calibrats es poden veure a la Figura 32 on, per ordre d'esquerra a dreta, es mostra: la mitjana dels errors MAE (error mitjà absolut), el millor error MAE obtingut, la mitjana de la diferència de morts entre l'esperat i l'obtingut, i la diferència més petita obtinguda.



predefinits en color groc, i els valors inferits per les dues calibracions en lila i vermell, anàlogament a la gràfica anterior. De la mateixa manera que abans, la segona calibració ens porta a uns valors més propers als predefinits amb anterioritat.



La millor calibrat trobat fins al moment, que ha estat representada en color vermell en els gràfics anteriors, ens dona el resultat mostrat a la següent gràfica: on la corba blava representa el nombre de morts estimat pel model i la corba taronja representa el nombre de morts real.

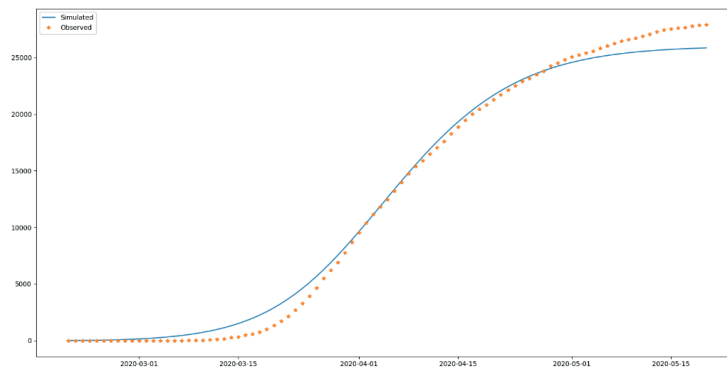


Figura 34. La millor calibrat trobat fins al moment amb el nombre de morts real.

A continuació veiem la distribució del nombre de casos actius per infectats asimptomàtics (asymptomatic, A) i simptomàtics (infected, I), i finats (dead, D).

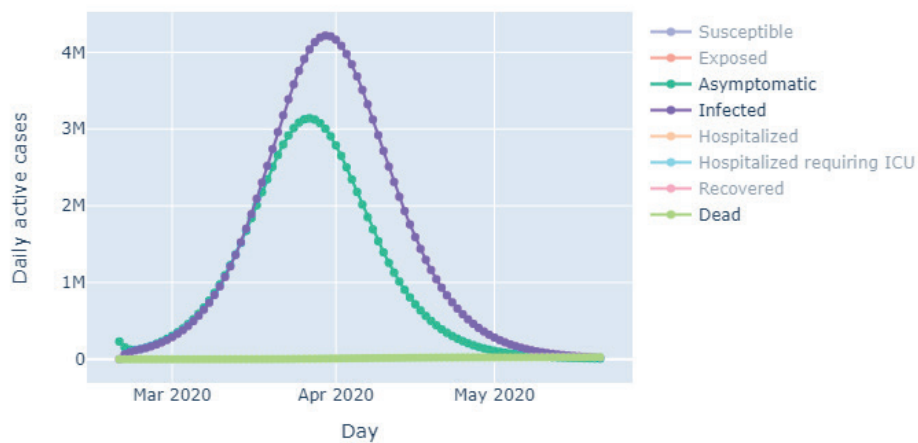


Figura 35. La distribució del nombre de casos actius per infectats asimptomàtics (asymptomatic, A) i simptomàtics (infected, I), i finats (dead, D)

La següent gràfica mostra el nombre de casos nous diaris per als compartiments que acabem d'esmentar. Es pot veure que diàriament es donen més casos sense símptomes que amb símptomes, però, aquest comportament és diferent quan parlem al nombre de casos actius. De forma activa hi ha menys casos sense símptomes que amb símptomes degut a la duració de la fase asimptomàtica, que és més curta.

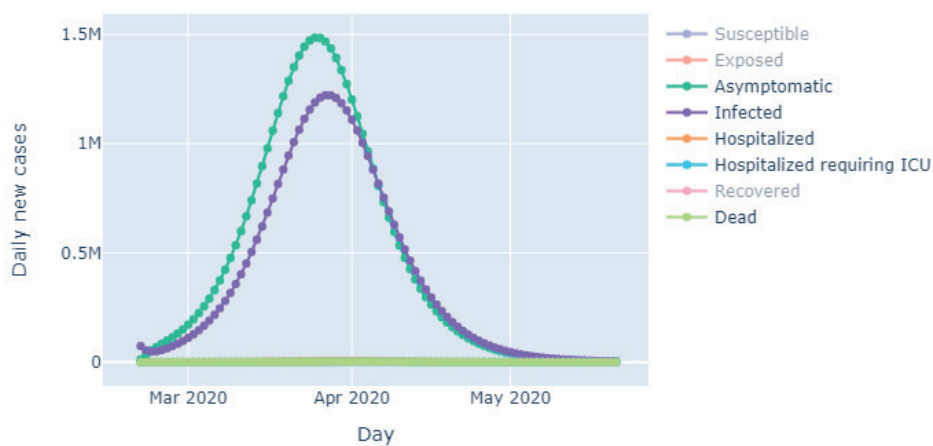


Figura 36. El nombre de casos nous diaris per als compartiments d'asimptomàtics, simptomàtics, hospitalitzats, ingressats a l'UCI, i finats.

A la següent gràfica es mostra també el nombre de casos nous diaris, en particular es mostren els casos de persones hospitalitzades fora (hospitalized, H) i dins de l'UCI (hospitalized requiring ICU, U) i de les persones finades (dead, D).

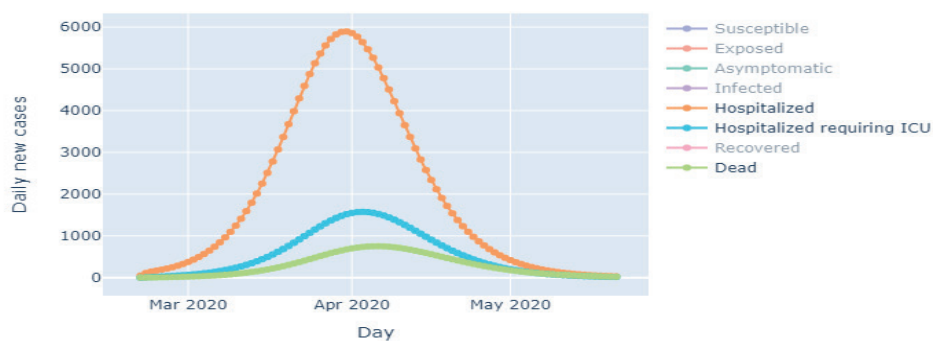


Figura 37. El nombre de casos nous diaris, de persones hospitalitzades fora (hospitalized, H) i dins de l'UCI (hospitalized requiring ICU, U) i de les persones finades (dead, D)

Donats els valors obtinguts pels paràmetres del model gràcies a aquest calibrat, podem executar el model sobre la població espanyola. Així, obtenim el comportament mostrat a la següent gràfica. Aquesta gràfica mostra els valors esperats i els valors obtinguts per a tres compartiments diferents: persones hospitalitzades (H), ingressades a l'UCI (U), i finades (D).

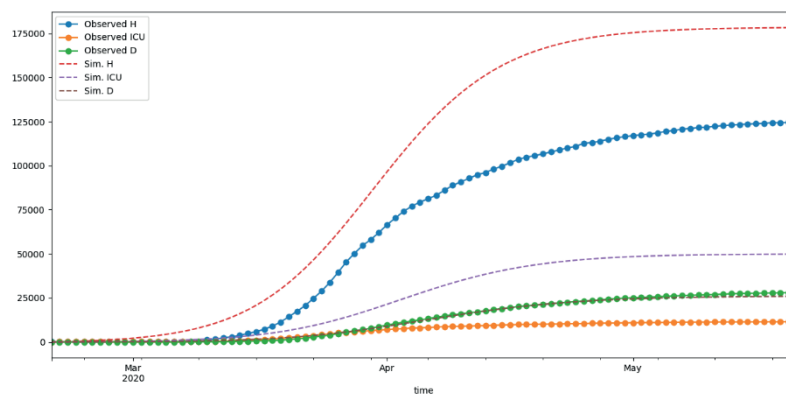


Figura 38. Els valors esperats i els valors obtinguts per a tres compartiments diferents: persones hospitalitzades (H), ingressades a l'UCI (U), i finades (D).

Simulacions d'escenaris pandèmics amb el model dirigit per dades

En aquesta primera simulació s'han recollit les dades d'infecció de Flow-Maps des del dia 19 de Gener de 2020. Les dades de la primera i segona onada s'han fet servir per a entrenar una xarxa LSTM amb 100 neurones. Les dades de la tercera onada s'han fet servir com a validació separant les dades dels últims 21 dies per a test. La xarxa s'ha entrenat per a minimitzar l'error quadràtic mig com a funció de cost i s'ha entrenat durant 500 èpoques al clúster de GPUs d'Eurecat.

A la següent gràfica mostrem els resultats de l'entrenament:

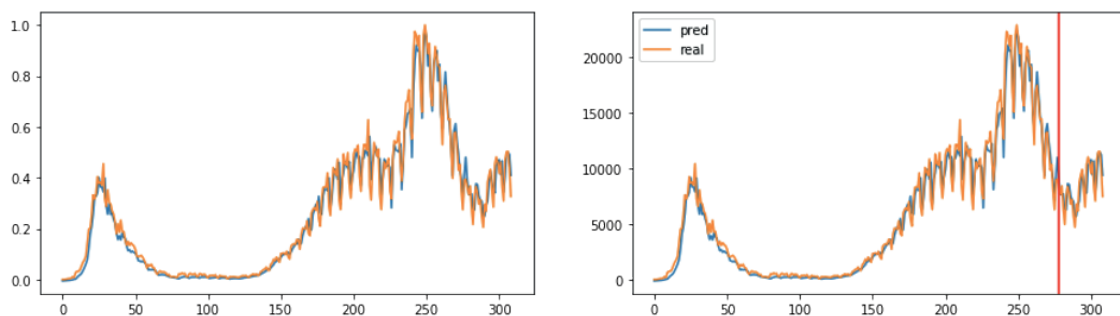


Figura 39. Resultats de l'entrenament

A la següent gràfica mostrem l'evolució prevista en casos acumulats pels propers 21 dies (en vermell)

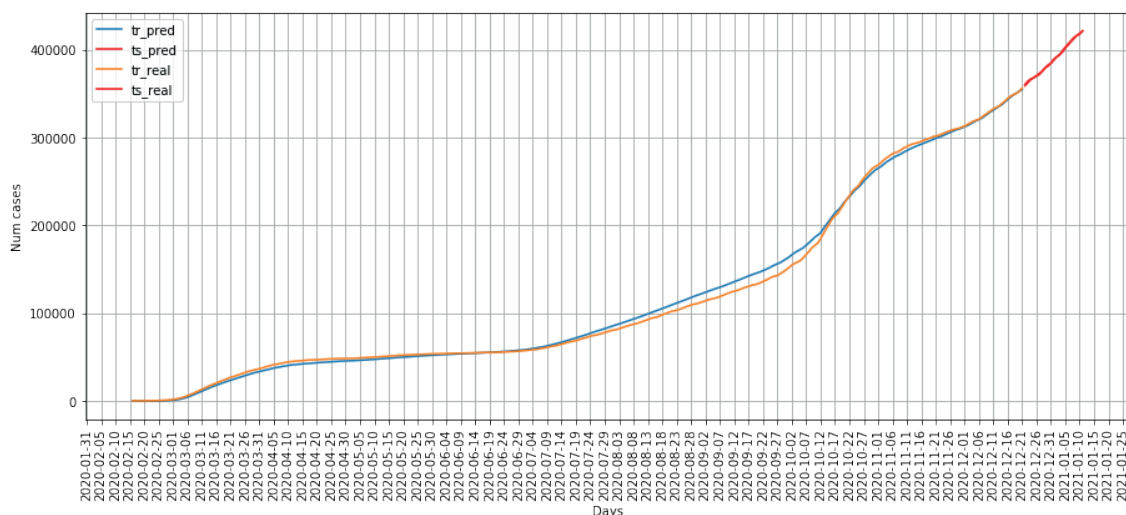


Figura 40. L'evolució prevista en casos acumulats pels propers 21 dies

L'error de predicció obtingut per la població de test és de 1531 casos. Independentment de l'error, la tendència de la corba és clarament creixent. Per aquest motiu, hem realitzat una predicció de l'evolució de la pandèmia en els propers dies començant el dia 18 de Gener de 2021 (últim dia de dades disponible) fins el dia 7 de Febrer de 2021. Així mateix, també hem estudiat la R_t (taxa d'infecció) durant aquest període. El següent gràfic mostra els resultats obtinguts de la primera predicció amb dades fins el dia 18/01/21.

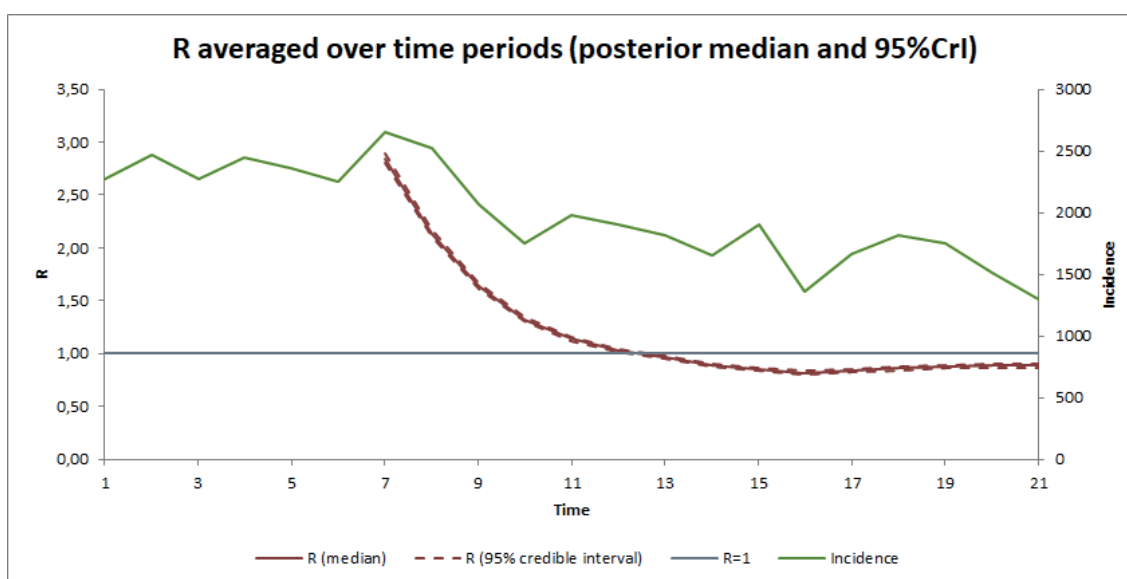


Figura 41. La R_t (taxa d'infecció) durant el dia 18 de Gener de 2021 (últim dia de dades disponible) fins al dia 7 de Febrer de 2021.

Tot seguit afegim la segona predicció realitzada amb dades fins el dia 22/01/21

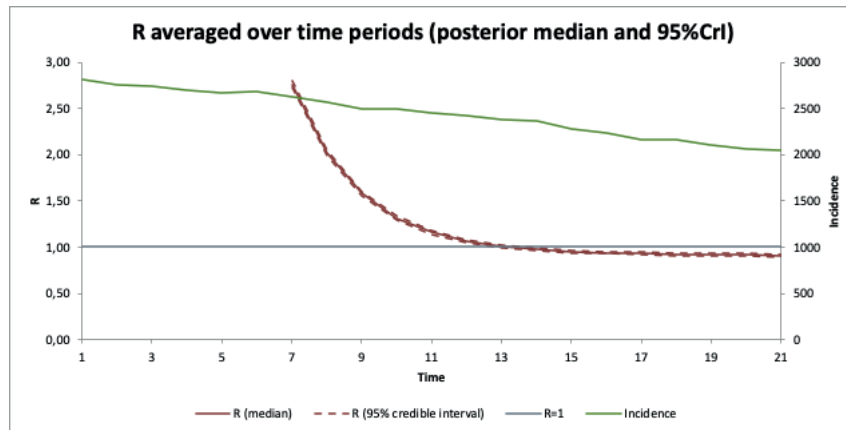


Figura 42. La R_t (taxa d'infecció) durant el dia 18 de Gener de 2021 (últim dia de dades disponible) fins el dia 7 de Febrer de 2021.

Malgrat la tendència del nombre de casos i de la R_t sigui a la baixa. Aquesta R_t s'estabilitza a 0,94. La interpretació d'aquest resultat és que la pandèmia seguirà amb un creixement exponencial fins a finals de Gener i s'estabilitzarà en un creixement lineal durant la primera setmana de Febrer. Aquest comportament es replica amb l'ocupació de llits de cures intensives (i.e. creixement exponencial fins al final de Gener i lineal durant Febrer). Cal comentar també que el model actual no pren en consideració la propagació de la soca britànica B117 al principat, que actualment s'estima que afecta un 5% dels infectats. Tot i això, l'augment que observem en la R_t estimada podria estar alineat amb l'augment de la taxa d'infecció que s'observa amb la soca britànica. L'augment sostingut d'aquesta soca implica un augment de la R_t en el temps podria acabar suposant un creixement exponencial en un horitzó temporal de tres setmanes.

La tendència actual en la R_t observada es correspon a la previsió de creixement observada en el nombre acumulat d'ingressos en UCI. A la següent gràfica mostrem la previsió dels ingressos acumulats en UCI fins el 12 de Febrer.

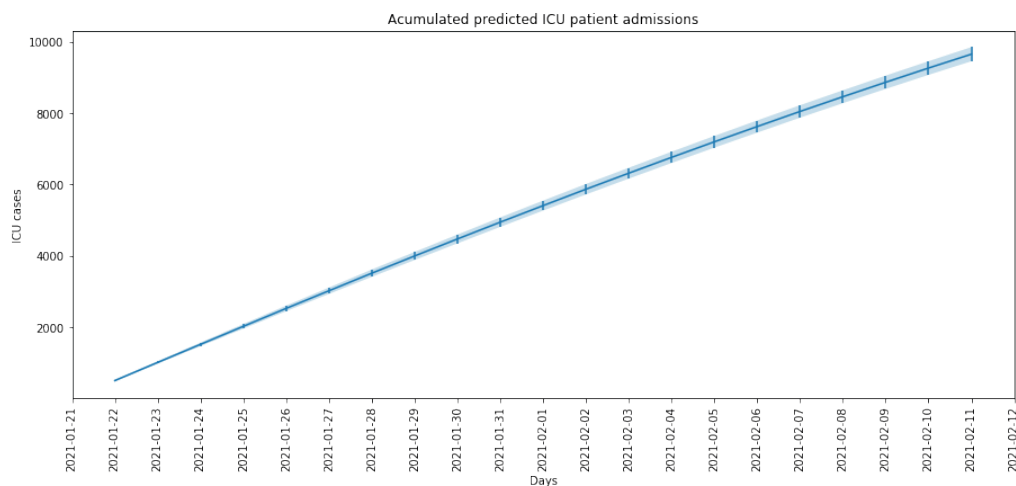


Figura 43. Nombre acumulat d'ingressos a l'UCI en les 10 prediccions realitzades.

Finalment a la següent gràfica mostrem la previsió diària d'ingressos a l'UCI. La mitjana prevista de casos diaris és de 447,6 $\pm$ 20,2 casos.

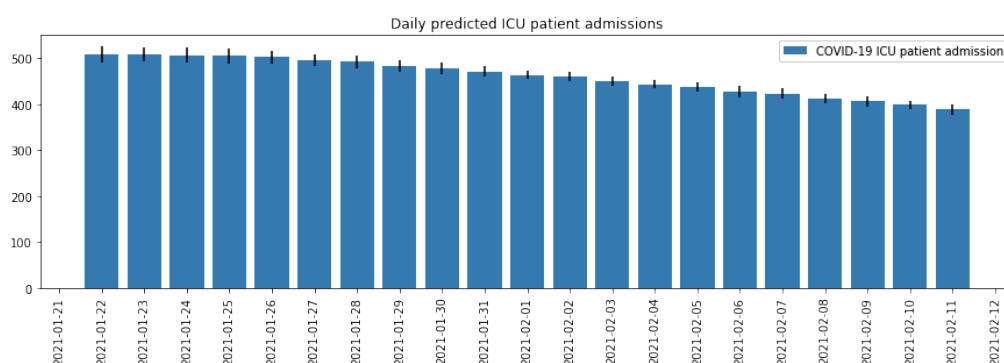


Figura 44. Previsió diària d'ingressos en UCI

Reflexions i recomanacions sobre les fonts de dades

Durant aquesta primera fase del projecte s'han emprat múltiples fonts de dades, que s'han detallat anteriorment. Amb la finalitat de poder realitzar el millor calibrat possible dels models i, també, per a poder estudiar diferents escenaris pandèmics es necessita disposar de dades el més acurades possible i amb un nivell de granularitat suficient per a identificar les diferents àrees bàsiques de salut (ABS). Aquest fet és especialment delicat pel que fa a les dades de mobilitat atès que les dades de l'INE venen determinades per les àrees de mobilitat. Una àrea de mobilitat pot correspondre a un municipi, però també pot donar-se el cas que una àrea de mobilitat inclogui diversos municipis. Això ha estat una limitació durant les primeres etapes del projecte fins que no s'ha disposat de les dades del Ministerio de Transportes (MITMA), que defineixen àrees de mobilitat a partir de les quals es pot fer un mapeig força acurat de les diferents ABS.

La necessitat de signar convenis amb els proveïdors de dades més importants (AQUAS i MITMA) i el compliment de la normativa de privacitat de dades ha suposat un retard considerable en el transcurs del projecte. Tampoc ha ajudat el fet de no poder accedir a dades de mobilitat d'operadors telefònics.

Pel que fa a les dades d'AQUAS, que s'han fet servir per a calibrar el model, procedeixen del IMDB i, per tant, arriben amb un retard considerable que fa inviable utilitzar aquesta font de dades per a fer prediccions sobre l'evolució de la pandèmia. Aquestes dades arriben un cop s'ha tancat el procediment de facturació dels hospitals i, per tant, arriben amb aproximadament tres mesos de retard de mitjana. A dia d'avui s'estan estudiant alternatives amb el departament de Salut i amb AQUAS per a agilitzar l'accés a les dades arribant a un compromís entre la qualitat i el temps en que les dades estan disponibles pel seu anàlisi.

Tot l'equip científic el projecte ha estat d'acord en que la millor aproximació consisteix en utilitzar les dades d'AQUAS per a fer els calibrats del model i explicar la dinàmica de la pandèmia i en fer servir les dades reportades de casos i defuncions al portal de dades obertes de la Generalitat per a poder fer prediccions sobre l'evolució de la pandèmia malgrat aquestes dades no siguin 100% acurades. Aquesta aproximació mixta ens ha de permetre arribar a un compromís per a poder fer prediccions i interpretar la dinàmica de la pandèmia còpsant-ne tota la seva complexitat.

Anteriorment hem comentat que un dels reptes més importants que ens trobem en l'estudi de la pandèmia consisteix en l'estimació del nombre de casos asimptomàtics. De fet, en tots els models que hem pogut estudiar incloent el nostre, la població d'asimptomàtics s'ha inclòs com a condició inicial al model. És a dir, s'ha calibrat com un paràmetre més. Aquest fet augmenta la complexitat dels mètodes de calibrat, que augmenta exponencialment amb cada paràmetre que s'hi ha d'estimar.

Un altre repte referent a les fonts de dades que s'ha hagut d'abordar durant el projecte ha estat la discrepància entre les diferents fonts així com els criteris de diagnòstic. Per exemple, durant la primera onada, el principal criteri de diagnòstic de la COVID-19 estava basat en símptomes i signes clínics i la confirmació per PCR era força limitada en molts casos. Un cop es va generalitzar l'ús de la PCR com a prova diagnòstica va augmentar significativament la positivitat de casos i va suposar un repte poder comparar la primera i la segona onada. També la inclusió de proves d'antígens ha suposat una adaptació dels models per a poder garantir la comparativa entre diferents onades de la pandèmia. A dia d'avui, l'arribada de la variant B117 del SARS-COV2 implica la competència entre dues soques del mateix virus de les quals una té una taxa de reproducció més alta que l'altra i un infectat transmet la malaltia a més persones en tots els rangs d'edat. Aquesta nova soca ha afegit una capa més de complexitat al model de manera que els calibrats i interpretacions del model i la seva dinàmica són encara més complexes. Amb la finalitat de poder extraure el màxim de profit del model epidemiològic creiem important destacar la necessitat d'augmentar els tests genètics amb la finalitat d'identificar diferents soques del virus per a poder estudiar la seva infectivitat així com l'efectivitat de les polítiques de restricció de mobilitat implementades.

L'arribada de les vacunes per la COVID-19 ha de suposar un avenç significatiu en el control de la pandèmia i la seva evolució. Per aquest motiu, també es fa necessari disposar de les taxes de vacunació per cada ABS i adaptar el model per a poder predir l'evolució de la malaltia.

El model que s'ha desenvolupat és prou escalable i modular per a incloure tota aquesta complexitat però la qualitat de les prediccions i escenaris estudiats serà tant bona com bones siguin les dades que s'han emprat per a calibrar-lo. Aquesta necessitat és encara més imperiosa per a l'aproximació dirigida per dades que pretenem encetar en una futura fase del projecte en la qual agilitzarem el procediment de predicció i d'estudi d'escenaris mitjançant tècniques d'aprenentatge automàtic amb xarxes neuronals. Les prediccions realitzades mitjançant aquestes tècniques s'explicaran amb el model epidemiològic que s'ha desenvolupat durant aquesta primera fase del projecte emprant les dades obertes de la Generalitat.

Conclusions i recomanacions finals



Aquest document presenta les activitats i resultats obtinguts durant la primera fase del projecte que vol esdevenir el futur Observatori Epidemiològic de Catalunya, o en clau més ambiciosa, l'Observatori estratègic de Dades de Salut.

S'han presentat les diferents fonts de dades utilitzades i els diferents models desenvolupats durant aquesta primera fase. De forma més específica, s'ha desenvolupat un model de mobilitat, un model mediambiental i dos models epidemiològics amb dades de mobilitat i d'infecció. Així mateix, també s'ha desenvolupat tota la metodologia per a que els models epidemiològics incloguin a futur d'altres fonts de dades com, per exemple, dades d'aigües residuals i altres intervencions com, per exemple, tractaments i campanyes de vacunació.

Pel que fa a les dades de mobilitat, s'ha confirmat la seva elevada correlació amb la transmissió de la pandèmia i la taxa d'infecció. També s'ha comprovat l'efecte de les restriccions en la mobilitat en la transmissió de la malaltia. Per aquest motiu, es considera una font de dades fonamental per al seguiment de la pandèmia i la predicció de la seva evolució. Un requisit indispensable per a poder continuar el projecte i fer prediccions és poder continuar disposant d'aquesta font de dades.

L'efecte de les variables meteorològiques va ser molt semblant al comportament de la mobilitat. A mesura que les condicions meteorològiques van ser més favorables, va haver-hi més mobilitat i per tant més contactes, el que va comportar un augment de la incidència.

L'associació entre les variables meteorològiques i les hospitalitzacions, els ingressos en UCI i les morts, és molt probable que hagi estat la conseqüència d'un augment de la incidència i no tant d'una possible relació directa entre elles.

L'exposició a la contaminació atmosfèrica, en particular per NO_2 , va estar associada a un augment de la incidència, encara que, sobre tot, durant la primera onada. Però, a més, aquesta associació només es va trobar en aquelles ABS més contaminades i en aquelles amb una contaminació moderada. Per tant, podria ser que l'exposició a la contaminació atmosfèrica fos un substitut d'altres variables que fossin les veritablement associades amb la incidència (baix nivell socioeconòmic, per exemple).

Igual que les variables meteorològiques, l'associació entre l'exposició a la contaminació atmosfèrica i l'agreuament de la malaltia, encara que en aquest cas no s'ha trobat una relació amb les morts, pot ser només una conseqüència de l'associació amb la incidència.

El primer model epidèmic que s'ha desenvolupat està basat amb equacions diferencials i s'ha calibrat amb dades de l'INE. Ja s'ha començat en el calibrat d'aquest model amb les dades d'AQUAS. L'objectiu principal d'aquest model consisteix en explicar la dinàmica de la pandèmia, valorar l'aparició de noves soques i l'efecte de diferents iniciatives com, per exemple, confinaments i campanyes de vacunació i/o tractament. Així mateix, aquest model es fa servir per a interpretar els resultats obtinguts amb el model dirigit per dades.

Malgrat no estava contemplat a l'abast d'aquesta primera fase del projecte, també s'ha desenvolupat un model dirigit per dades per a poder fer les primeres prediccions de l'evolució de la pandèmia arreu de Catalunya. S'han presentat els resultats de les dues prediccions realitzades amb un horitzó de 21 dies realitzades a dia 18/01/22 i a 22/01/22. Les dues prediccions mostren una estabilització de la R_t (taxa d'infecció estimada com el nombre de persones que un positiu infecta en mitjana) respectivament a 0,88 i a 0,94. Aquesta taxa d'infecció es reflexa també en la predicció del nombre d'ingressos en unitats de cures intensives, que creix de forma lineal amb, aproxi-

madament, 470 ingressos diaris de mitjana al principat.

Per aquest motiu, considerem necessari seguir monitoritzant periòdicament l'evolució de la pandèmia amb aquest model i actualitzar les prediccions realitzades a mesura que es vagin actualitzant les dades al portal de dades obertes de la Generalitat.

Per a la segona fase del projecte es proposa continuar amb les actualitzacions de les prediccions, completar el sistema de predicció afinant les prediccions d'ingressos hospitalaris i en unitats de cures intensives a tot el territori i per àrea bàsica de salut amb la finalitat de poder definir aquelles àrees que necessitaran reforçar-se. També es proposa millorar tots els models desenvolupats en aquesta fase incloent-hi noves fonts de dades com, per exemple, dades d'aigües residuals en les quals s'hi analitza la presència del virus SARS-COV2 i les seves variants. Sobre tot, el model dirigit per dades pel seu gran poder predictiu i eficiència. Aquest model ens fa diferencials respecte altres iniciatives. Així mateix, també es proposa continuar treballant en els models basats en equacions diferencials per a interpretar els resultats obtinguts amb el model dirigit per dades.

Un lliurable principal d'aquesta segona fase serà un quadre de comandament de visualització de dades amb accés personalitzat segons els rols dels diferents usuaris finals. Aquest quadre de comandament també inclourà un sistema alertes identificant, per exemple, aquelles zones on es preveu un augment significatiu de casos.

Finalment, també es busca consolidar el gran equip multidisciplinari que s'ha establert al projecte amb membres amb molt coneixement específic i de gran valor afegit (e.g. ISGlobal i Hospital Clínic).



Annexos



Annex I: Origen de les dades del model de mobilitat

Taula 3. Descripció dels principals orígens de dades que s'empren per alimentar el sistema de Flow-Maps.

Orígen	URL
Catalunya	Mapes https://salutweb.gencat.cat/ca/el_departament/estadistiques_sanitaries/cartografia/ https://www.icgc.cat/Administracio-i-empresa/Descarregues/Capes-de-geoinformacio/Base-municipal Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i municipi https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/jj6z-iyrp Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i àrea bàsica de salut https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/xuwf-dxjd Registre central de població del CatSalut https://analisi.transparenciacatalunya.cat/Salut/Registre-central-de-poblaci-del-CatSalut/ftq4-h9vk Registre de defuncions per COVID-19 a Catalunya. Segregació per sexe i comarca https://analisi.transparenciacatalunya.cat/Salut/Registre-de-defuncions-per-COVID-19-a-Catalunya-Se/uqk7-bf9s
Madrid	Mapes https://www.madrid.org/nomecalles/DescargaBDTCorte.icm https://datos.comunidad.madrid/catalogo/dataset/7da43feb-8d4d-47e0-abd5-3d022d29d09e/resource/04f71acb-e0c1-47be-84c9-1f57d2e09e40/download/municipios_y_distritos_madrid.zip Covid 19 -TIA per Municipis i Districtes de Madrid http://datos.comunidad.madrid/catalogo/dataset/covid19_tia_muni_y_distritos https://datos.gob.es/es/catalogo/a13002908-covid-19-tia-por-municipios-y-distritos-de-madrid1 Covid 19 -TIA Zones Bàsiques de Salut https://datos.comunidad.madrid/catalogo/dataset/covid19_tia_zonas_basicas_salud https://datos.gob.es/es/catalogo/a13002908-covid-19-tia-zonas-basicas-de-salud1

Orígen	URL
Euskadi	Mapes https://www.geo.euskadi.eus/zonas-de-salud-de-euskadi/s69-geodir/es/ https://www.geo.euskadi.eus/s69-geoser/eu/contenidos/informacion/servicio ftp/eu 80/servicio ftp.html ftp://ftp.geo.euskadi.eus/cartografia Evolució del coronavirus (COVID-19) en Euskadi https://opendata.euskadi.eus/catalogo/-/evolucion-del-coronavirus-covid-19-en-euskadi/
INE	Mapes, metadades i dades https://www.ine.es/covid/covid_movilidad.htm Mapes https://www.ine.es/ss/Satellite?L=es_ES&c=Page&cid=1259952026632&p=1259952026632&pagename=ProductosYServicios%2FPYSLayout
Navarra	Mapes https://gobiernoabierto.navarra.es/es/open-data/datos/servicio-navarro-salud-delimitacion-zonas-basicas-salud https://gobiernoabierto.navarra.es/es/open-data/datos/servicio-navarro-salud-delimitacion-areas-salud Dades zones bàsiques de salut, covid-19 https://gobiernoabierto.navarra.es/es/open-data/datos/datos-zonas-basicas-salud-covid-19 Dadets població zones bàsiques de salut, covid-19 https://gobiernoabierto.navarra.es/es/open-data/datos/datos-poblacion-zonas-basicas-salud-covid-19
Castilla i Lleó	Mapes https://datosabiertos.jcyl.es/web/jcyl/set/es/medio-ambiente/zonas-basicas-salud-cyl/1284688135689 Taxa de malalts per zones bàsiques de salut https://datosabiertos.jcyl.es/web/jcyl/set/es/salud/tasa-coronavirus-zonas-basicas-salud/1284942912395 Situació epidemiològica coronavirus (COVID-19) a Castilla i Lleó per hospitals https://datosabiertos.jcyl.es/web/jcyl/set/es/salud/situacion-coronavirus-hospitales/1284941728695

Orígen	URL
Aragó	Publicacions i recursos relacionats amb el Coronavirus a Aragó https://opendata.aragon.es/datos/catalogo/dataset/publicaciones-y-anuncios-relacionados-con-el-coronavirus-en-aragon https://datos.gob.es/es/catalogo/a02002834-publicaciones-y-recursos-relacionados-con-el-coronavirus-en-aragon1
CNE-COVID	Panell COVID-19 per províncies i comunitats autònomes https://cnecovid.isciii.es/covid19/
Cantabria	Dades https://www.scsalud.es/coronavirus
MoMo	Panell MoMo (excés de mortalitat) per comunitats autònomes https://momo.isciii.es/public/momo/dashboard/momo_dashboard.html
CNIG	Mapes http://centrodedescargas.cnig.es/CentroDescargas/
València	COVID-19 Casos confirmats per PCR +, casos PCR + en els últims 14 dies i persones mortes per municipi de la Comunitat Valenciana 2020 https://dadesobertes.gva.es/dataset/covid-19-casos-confirmats-pcr-casos-pcr-en-els-ultims-14-dies-i-persones-mortes-per-municipio-2020

Taula 4. Descripció de les principals capes geogràfiques

Font	ID: Descripció
	ine_mov: Àrees de mobilitat INE ine_sec: Seccions censals INE
MITMA	mitma_mov: Àrees de mobilitat MITMA mitma_mun_mov: Àrees de mobilitat amb municipis MITMA
CNIG	cnig_municipios: Municipis CNIG cnig_provincias: Províncies CNIG cnig_ccaa: Comunitats Autònomes CNIG
Catalunya	reg_san_09: Regions Sanitàries GenCat sect_san_09: Sectors Sanitaris GenCat aga_09: Àrees de Gestió Assitencial GenCat abs_09: Àrees Bàsiques de Salut GenCat shires_09: Comarques Catalunya
Madrid	mun_dis_13: Municipi+districte de Madrid zon_bas_13: Zones bàsiques sanitàries de Madrid area_san_13: Àrees de Salut C. Madrid dis_san_13: Districtes de Salut C. Madrid zon_san_13: Zones de Salut C. Madrid zbs_13: Zones Bàsiques de Salut C. Madrid
Euskadi	oe_16: Osasun Eremuak (Zones de Salut) Euskadi cs_hosp_16: Centres de Salut, ambulatoris i hospitals públics d'Euskadi
Navarra	zbs_15: Zones Bàsiques de Salut del Servei Navarro de Salut as_15: Àrees de Salut del Servei Navarro de Salut
Castella i Lleó	zbs_07: Zoneas Bàsiques de Salut de CiL
	comarcas_12: Comarques de Galícia concellos_12: Municipis de Galícia parroquias_12: Comarques de Galícia
	dis_san_01: Districtes sanitaris d'Andalusia zbs_01: Zones bàsiques de Salut d'Andalusia d_zbs_01: Demarcacions dins de Zones bàsiques de Salut d'Andalusia
Otras	world: World map ate_pri: Centres d'atenció primària ate_urg_extra: Centres d'atenció urgent extrahospitalària hospitales: Centres d'atenció hospitalària codigos_postales: Codis Postals andorra: Andorra

Annex II: Model Ambiental en detall

Població i període d'estudi

La població d'estudi va estar constituïda per 373 Àrees Bàsiques de Salut (ABS) de Catalunya. El període d'estudi va ser el comprés entre el 28 de febrer i el 31 d'agost de

2020. El disseny no va ser equilibrat, ja que no en totes les ABS es van observar casos de les variables resultat el mateix número de dies.

Variables resultat

Es van considerar, com a variables resultat, i per a cada ABS, el número diari de casos identificats com a positius per PCR (com a variable que aproximava la incidència); el número diari d'hospitalitzacions per COVID-19; el número diari d'ingressats a les UCI com a conseqüència del COVID-19; i el número diari de morts.

D'acord al registre RSAcovid19 del Departament de Salut, les morts s'obtenen a partir de les declaracions de les empreses de serveis funeraris. Només es tenen en compte les defuncions relacionades amb la COVID-19: que el motiu del certificat de defunció especifiqui que la persona ha mort a causa de la COVID-19, tingui o no un prova diagnòstica PCR o test d'antígens amb resultat positiu (Govern Obert, Generalitat de Catalunya, 2020).

Variables meteorològiques

A més de les variables meteorològiques indicades més amunt, i amb l'objectiu de recollir les possibles interaccions entre elles, es va construir un índex de confort tèrmic. El confort tèrmic és subjectiu i es defineix com la sensació de satisfacció amb l'entorn tèrmic. D'entre els possibles índexs, vam escollir la temperatura efectiva

neta, la qual estima com la temperatura, la humitat i la velocitat del vent afecten a la sensació de confort (Blazczyk et al., 2012; Li i Chan, 2000; Davis et al., 2016). La temperatura efectiva neta s'ha trobat relacionada amb algunes malalties, en concret les cardiovasculars (Santurín et al., 2020).

Contaminants atmosfèrics

És important tenir en compte que hem avaluat l'exposició a llarg termini als contaminants atmosfèrics. És a dir, una persona en virtut de residir en una determinada ABS, va

estar exposat a un nivell (mitjà) de diversos contaminants atmosfèrics a l'ABS on, en el nostre cas, va residir almenys durant el període 2011- 2019.

Estimació dels nivells de les variables meteorològiques i dels contaminants atmosfèrics

La distribució espacial tant de les estacions meteorològiques com dels punts de mesurament de la contaminació atmosfèrica ni és homogènia ni coincideix amb la distribució de les ABS en el territori de Catalunya. Hi ha moltes ABS que no tenen cap estació meteorològica i/o punt de

mesurament de la contaminació. Per altra banda, hi ha algunes que tenen varies estacions i/o punts.

Per aquest motiu, vam haver d'estimar els valors de les variables meteorològiques i els nivells de contaminants

atmosfèrics als que van estar exposats els residents en cadascuna de les ABS durant l'any 2020 (per a les variables meteorològiques) i des del 2011 al 2019 (pels contaminants atmosfèrics).

Per fer-ho vam utilitzar un model Bayesià conjunt en el que la variable dependent era cadascuna de les variables meteorològiques (temperatura, humitat relativa, pressió atmosfèrica, velocitat del vent i irradiació solar) i cadascun dels contaminants (PM10, NO2 i O3). Com a variables explicatives van incloure la resta de variables meteorològiques i de contaminants (llevat de la dependent), respectivament. En el cas dels contaminants, vam

incloure, a més, la resta de contaminants explicats més a dalt (PM2.5, NO, SO2, CO, C6H6, B, plom, arsènic, níquel i cadmi). En tots els casos vam controlar per la dependència espacial (el fet que els valors de les variables meteorològiques i dels nivells dels contaminants d'ABS properes siguin més semblants que els de les ABS més allunyades) i temporal (el fet que els valors i nivells d'un dia determinat siguin molt més semblants a la dels dies immediatament anteriors que a la resta de dies), així com per la superfície de l'ABS (quanta més extensa sigui l'ABS menys dependència espacial i/o temporal presentaran aquests valors i nivells). Els detalls del model es poden consultar en Barceló et al. (2016).

Variables socioeconòmiques

A excepció de la densitat de població, totes les altres variables socioeconòmiques es van observar a nivell de secció censal. Utilitzant la població de cadascuna de les seccions censals com a ponderacions, vam calcular la

mitjana ponderada dels valors d'aquestes variables a les seccions censals que va compondre l'ABS, per obtenir el seu valor a nivell d'ABS. La font de la població de la secció censal, total i per sexe, va ser l'INE (2020b).

Models d'associació

Per avaluar l'associació entre les variables resultat i les variables mediambientals (meteorològiques i contaminants) van utilitzar models lineals generalitzats mixtes amb una funció vincle de Poisson. Aquests són equivalents a models de regressió, en els que la variable dependent és una variable discreta (en el nostre cas, el número de casos diaris de cadascuna de les variables resultat).

Per a cadascuna de les variables resultat vam estimar sis models. En primer lloc, vam estimar models per a tot el període (28 de febrer-31 d'agost), i models per a cada un de dos subperíodes: primera onada (28 de febrer-30 de juny) i resta del període (1 de juliol-31 d'agost), període vall a partir d'ara. En segon lloc, tant per tot el període com pels dos subperíodes vam estimar dos models, un que incloïa la temperatura efectiva neta amb la resta de contaminants i de variables meteorològiques (excepte les variables que la componen, és a dir, temperatura, humitat i velocitat del vent); l'altra que incloïa tots els contaminants i totes les variables meteorològiques (excepte la temperatura efectiva neta).

En els models vam controlar per la mobilitat, aproximat pel percentatge de persones que es mouen en l'ABS; i pel percentatge de persones que es mouen fora de la mateixa àrea, respecte a les que es mouen en l'ABS; per a les variables socioeconòmiques: renda mitjana per persona en l'ABS; taxa d'atur en l'ABS; densitat de la població en l'ABS; percentatge d'infrahabitatges en l'ABS; percentatge de persones de 65 anys o més en l'ABS; percentatge d'estrangers de països de baixa renda; pel confinament (en el model per a tot el període); així com per a la població de cada ABS.

Tanmateix, vam controlar per la dependència espacial (és a dir, per la distribució espacial de la pandèmia) i per la dependència temporal (és a dir, per la dinàmica temporal de la pandèmia). Amb aquest fi i en aquest cas, vam utilitzar efectes aleatoris (els detalls dels models es poden consultar en Saez et al., 2020).

Preveient que existís un retard entre la infecció i la diagnòsi o entre aquesta i l'agreujament de la malaltia (aproximada per les hospitalitzacions, els ingressos en

UCI i la mort), tant les variables meteorològiques com les de mobilitat van ser introduïdes en el model com una mitjana mòbil de 10 dies (el dia corresponent i els 9 dies anteriors). Vam permetre que les relacions entre les variables resposta i les variables meteorològiques i els con-

taminants, així com també la renda mitjana per persona, la taxa d'atur i la densitat d'habitatges fos en realitat no lineal. Per aconseguir-ho vam categoritzar totes aquestes variables en quartils.

Resultats

Els resultats dels models per a tot el període difereixen, en algun cas molt, dels models per als dos subperíodes, suggerint que, en totes les variables resultat, van existir dos patrons diferents que no es poden modelitzar conjuntament i, per tant, s'haurien d'interpretar de forma separada.

Per altra banda, no s'haurien d'interpretar de forma separada les associacions entre les variables resposta i variables com temperatura, la humitat relativa i la velocitat del vent, doncs aquestes presenten evidents interaccions.

Pel que fa a la incidència (aproximada pels casos positius per PCR), el comportament de la temperatura neta efectiva va ser similar tant en la primera onada com en el període vall (Taula 5 i Figura 45). Quant més alta va ser la temperatura neta efectiva, major va ser el risc d'infecció. Pel que fa a les hospitalitzacions, en la primera onada, a mesura que augmenta la temperatura neta efectiva, també augmenten les hospitalitzacions (Taula 7 i Figura 47). Això també passa pel període vall, però amb una magnitud més gran que en la primera. En el cas dels ingressos en UCI, la temperatura neta efectiva només va ser estadísticament significativa per la primera onada (Taula 9 i Figura 49). El mateix comportament que les hospitalitzacions el veiem quan analitzem les morts, sent la diferència de riscos entre el primer i el segon subperíode molt més gran ara que en les hospitalitzacions (Taula 11 i Figura 51). En el període vall el risc de morir en augmentar la temperatura és molt més gran que en la primera onada.

En analitzar la pressió atmosfèrica veiem un comportament diferent tant pel que fa als dos subperíodes com pel que fa a les variables resposta. Quan va ser estadísticament significativa, la pressió va actuar com a factor protector en la primera onada i com a un factor de risc en el període vall. Però, la significació estadística va variar entre les variables resposta i inclús entre quartils.

Respecte a la irradiació solar i pel que fa a la variable positius PCR, no existeix tampoc un comportament diferent respecte als dos subperíodes (Taula 5 i Figura 45). En general, podem dir que a més irradiació solar, menor és el risc d'infecció (explicat per positius PCR). Aquesta variable no és estadísticament significativa a l'hora d'explicar els ingressos en UCI (Taula 9 i Figura 49). En canvi, pel que fa al nombre d'hospitalitzacions i morts (Taula 7 i Taula 11, i Figura 47 i Figura 51), hem d'analitzar els dos subperíodes per separat, ja que el seu comportament és diferent entre elles. Mentre que per la primera onada veiem que al augmentar la irradiació, disminueixen les hospitalitzacions (i morts), pel període vall, el comportament és diferent. En aquest cas, en augmentar la irradiació solar global, augmenten, en termes generals, les hospitalitzacions (i les morts, en major magnitud). És a dir, la irradiació solar passa a ser un factor de risc.

Pel que fa als contaminants atmosfèrics NO₂ i O₃ i el seu efecte sobre el risc d'infecció (aproximat per positius per PCR) (Taula 6 i Figura 46), trobem un comportament similar al que observàvem per la variable temperatura neta efectiva, ja que quan analitzem NO₂, no hi ha diferències estadísticament significatives entre els dos subperíodes i en augmentar el nivell de NO₂, el risc d'infecció augmenta. Si observem el comportament de O₃, aquest tampoc és diferent entre els dos subperíodes. En general, podem dir que el risc d'infectar-se és major per aquelles persones que resideixen en ABS amb uns nivells de NO₂ entre 52,96 i 60,37, respecte a aquelles que resideixen en ABS amb uns nivells de NO₂ menors a 45,98. Ara bé, respecte a PM₁₀, el comportament és diferent pels dos subperíodes considerats. Només pel període vall, podem dir que, en termes generals, en augmentar el nivell de PM₁₀, també augmenta el risc d'infecció.

Pel que fa als nombre d'hospitalitzacions i als ingressos en UCI, l'únic contaminant atmosfèric estadísticament significatiu és NO₂ (Taula 8 i Taula 10, i Figura 48 i Figura 50). En termes generals, podem dir, que tant en la primera onada com en el període vall, en augmentar el nivells de

NO₂, augmenten tant el número d'hospitalitzacions, com els ingressos en UCI. En canvi, si observem el número de morts, no hi ha cap contaminant atmosfèric que tingui cap efecte a l'hora d'explicar el comportament d'aquesta variable (Taula 12 i Figura 52).

Conclusions

L'efecte de les variables meteorològiques va ser molt semblant al comportament de la mobilitat. A mesura que les condicions meteorològiques van ser més favorables, va haver-hi més mobilitat i per tant més contactes, el que va comportar un augment de la incidència.

L'associació entre les variables meteorològiques i les hospitalitzacions, els ingressos en UCI i les morts, és molt probable que hagi estat la conseqüència d'un augment de la incidència i no tant d'una possible relació directa entre elles.

L'exposició a la contaminació atmosfèrica, en particular per NO₂, va estar associada a un augment de la incidèn-

cia, encara que, sobre tot, durant la primera onada. Però, a més, aquesta associació només es va trobar en aquelles ABS més contaminades i en aquelles amb una contaminació moderada. Per tant, podria ser que l'exposició a la contaminació atmosfèrica fos un substitut d'altres variables que fossin les veritablement associades amb la incidència (baix nivell socioeconòmic, per exemple).

Igual que les variables meteorològiques, l'associació entre l'exposició a la contaminació atmosfèrica i l'agreujament de la malaltia, encara que en aquest cas no s'ha trobat una relació amb les morts, pot ser només una conseqüència de l'associació amb la incidència.

Referències

Barceló MA, Varga D, Tobías A, Díaz J, Linares C, Saez M. Long term effects of traffic noise on mortality in the city of Barcelona, 2004-2007. *Environ Res* 2016; 147:193-206.

Blazejczyk K, Epstein Y, Jendritzky G, Staiger H, Tinz B. Comparison of UTCI to selected thermal indices. *Int. J. Biometeorol.* 2012; 56(3):515–535.

Davis RE, McGregor GR, Enfield KB. Humidity: a review and primer on atmospheric moisture and human health. *Environ. Res.* 2016; 144(Pt A):106-116.

Govern Obert. Generalitat de Catalunya. Dades obertes i COVID-19 [Disponible en: http://governobert.gencat.cat/en/dades_obertes/dades-obertes-covid-19/index.html], últim accés el 21 de gener de 2021].

INE. Instituto Nacional de Estadística. Indicadores para Secciones Censales. Censo Español de Población y Viviendas 2011 [Disponible en: http://www.ine.es/censos2011_datos/cen11_datos_resultados_seccen.htm, últim accés el 21 de gener de 2021].

INE. Instituto Nacional de Estadística. Atlas de distribución de renta de los hogares 2020a [Disponible en: https://www.ine.es/en/experimental/atlas/exp_atlas_tab_en.htm, últim accés el 21 de gener de 2021].

INE. Instituto Nacional de Estadística. Estadística del Padrón Continuo 2020b. [Disponible en: https://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736177012&menu=resultados&sec-c=1254736195461&idp=1254734710990#!tabs-1254736195557, últim accés el 21 de gener de 2021].

Li PW, Chan ST. Application of a weather stress index for alerting the public to stressful weather in Hong Kong. *Meteorol Appl.* 2000; 7:369-375.

PNUD. United Nations Development Program. Human Development Report 2019 [Disponible en: <http://hdr.undp.org/en/content/human-development-report-2019-readers-guide>, últim accés el 21 de gener de 2021].

Saez M, Tobías A, Barceló MA. Effects of long-term exposure to air pollutants on the spatial spread of COVID-19 in Catalonia, Spain. *Environ Res.* 2020; 191:110177.

Santurtún A, Almendra R, Fdez-Arroyabe P, Sánchez-Lorenzo A, Royé D, Zarrabeitia MT, Santana P. Predictive value of three thermal comfort indices in low temperatures on cardiovascular morbidity in the Iberian peninsula. *Sci Total Environ.* 2020; 729:138969.

Taula 5. Associació entre les variables meteorològiques i els casos positius per PCR.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
Temperatura (°C) [<8,61 °C]			
8,61 °C – 15,58 °C	0,986 (0,955, 1,018)		0,950 (0,925, 0,976)
15,58 °C – 23,50 °C	1,007 (0,947, 1,071)	1,170 (1,059, 1,291)	0,979 (0,934, 1,027)
> 23,50 °C	2,229 (1,524, 3,169)	1,109 (0,999, 1,230)	0,962 (0,910, 1,017)
Humitat relativa (%) [<61,82%]			
61,82% - 70,65%	1,069 (1,026, 1,115)	0,960 (0,937, 0,985)	0,914 (0,896, 0,932)
70,65% - 79,37%	1,144 (1,095, 1,195)	0,908 (0,876, 0,941)	0,915 (0,893, 0,936)
> 79,37%	1,115 (1,064, 1,169)	0,874 (0,807, 0,945)	0,908 (0,882, 0,935)
Velocitat del vent a 2 m (m/s) [<0,56 m/s]			
0,56 m/s – 0,76 m/s	0,936 (0,915, 0,957)	0,949 (0,920, 0,978)	0,946 (0,929, 0,963)
0,76 m/s – 1,08 m/s	0,993 (0,965, 1,021)	0,963 (0,932, 0,994)	0,989 (0,970, 1,009)
> 1,08 m/s	0,990 (0,958, 1,024)	1,022 (0,983, 1,062)	0,984 (0,961, 1,007)
Temperatura neta efectiva (°C) [<5,02 °C]			
5,02 °C – 12,46 °C	1,043 (1,013, 1,074)		1,031 (1,005, 1,058)
12,46 °C – 19,97 °C	1,157 (1,089, 1,230)	1,206 (1,101, 1,321)	1,149 (1,096, 1,204)
> 19,97 °C	1,663 (1,064, 2,492)	1,173 (1,066, 1,291)	1,158 (1,097, 1,222)
Pressió atmosfèrica (hPa) [<904,15 hPa]			
904,15 hPa – 955,64 hPa	0,960 (0,927, 0,994)	1,017 (0,934, 1,107)	1,029 (1,001, 1,058)
955,64 hPa – 993,98 hPa	1,008 (0,961, 1,058)	1,078 (0,989, 1,175)	1,064 (1,030, 1,098)
> 993,98	1,001 (0,949, 1,055)	1,057 (0,965, 1,158)	1,039 (1,001, 1,078)
Irradiació solar global (W/m²) [<101,54 W/m²]			
101,54 W/m² – 171,40 W/m²	0,956 (0,934, 0,978)	0,871 (0,843, 0,899)	0,932 (0,915, 0,949)
171,40 W/m² – 294,51 W/m²	0,907 (0,878, 0,937)	0,989 (0,953, 1,025)	0,944 (0,922, 0,966)
> 294,51 W/m²	0,890 (0,841, 0,941)	0,932 (0,886, 0,980)	0,901 (0,871, 0,932)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat.

En tots els casos, promig dels retards 1 a 10 de la variable.

Taula 6. Associació entre contaminants atmosfèrics i els casos positius per PCR.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
PM₁₀ (μ/m³) [<20,80 μ/m³]			
20,80 μ/m³ – 23,34 μ/m³	0,949 (0,845, 1,066)	1,153 (1,019, 1,303)	1,099 (0,985, 1,226)
23,34 μ/m³ – 25,89 μ/m³	0,935 (0,821, 1,065)	1,051 (0,916, 1,205)	1,062 (0,940, 1,201)
> 25,89 μ/m³	0,917 (0,787, 1,067)	1,290 (1,098, 1,516)	1,157 (1,002, 1,336)
NO₂ (μ/m³) [<21,93 μ/m³]			
21,93 μ/m³ - 25,77 μ/m³	1,136 (0,964, 1,338)	1,016 (0,896, 1,154)	1,080 (0,965, 1,210)
25,77 μ/m³ – 30,65 μ/m³	1,074 (0,943, 1,224)	0,915 (0,799, 1,049)	1,016 (0,899, 1,148)
> 30,64 μ/m³	1,172 (1,039, 1,322)	1,074 (0,904, 1,275)	1,122 (1,015, 1,230)
O₃ (μ/m³) [<45,98 μ/m³]			
45,98 μ/m³– 52,96 μ/m³	0,931 (0,821, 1,055)	0,918 (0,802, 1,050)	1,004 (0,892, 1,131)
52,96 μ/m³ – 60,37 μ/m³	1,109 (0,975, 1,260)	1,075 (0,943, 1,225)	1,111 (0,983, 1,254)
> 60,37 μ/m³	1,018 (0,887, 1,168)	0,997 (0,862, 1,152)	1,031 (0,906, 1,174)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat .

Ajustat per mobilitat (percentatge de persones que es mouen en l'ABS; i percentatge de persones que es mouen fora de la mateixa àrea, respecte a les que es mouen en l'ABS) (promig dels retards 1 a 10 i quartils); renda mitjana per persona en l'ABS (quartils); taxa d'atur en l'ABS (quartils); densitat de la població en l'ABS (quartils); percentatge d'infrahabitatges en l'ABS (quartils); percentatge de persones de 65 anys o més en l'ABS; percentatge d'estrangers de països de baixa renda (quartils); i confinament (en el model per a tot el període).

Taula 7. Associació entre les variables meteorològiques i els casos hospitalitzats.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
Temperatura (°C) [<8,61 °C]			
8,61 °C – 15,58 °C	1,005 (0,963, 1,049)		0,968 (0,930, 1,007)
15,58 °C – 23,50 °C	1,104 (1,007, 1,211)	1,569 (1,179, 2,082)	1,131 (1,045, 1,225)
> 23,50 °C	0,327 (0,003, 1,757)	1,346 (0,994, 1,818)	1,107 (0,995, 1,232)
Humitat relativa (%) [<61,82%]			
61,82% - 70,65%	1,095 (1,037, 1,157)	0,920 (0,853, 0,992)	0,922 (0,883, 0,961)
70,65% - 79,37%	1,193 (1,128, 1,263)	0,907 (0,816, 1,008)	0,992 (0,948, 1,038)
> 79,37%	1,190 (1,118, 1,266)	1,066 (0,854, 1,321)	0,991 (0,941, 1,043)
Velocitat del vent a 2 m (m/s) [<0,56 m/s]			
0,56 m/s – 0,76 m/s	0,915 (0,889, 0,942)	0,955 (0,869, 1,049)	0,916 (0,891, 0,942)
0,76 m/s – 1,08 m/s	0,972 (0,937, 1,009)	0,949 (0,861, 1,047)	0,969 (0,937, 1,002)
> 1,08 m/s	0,957 (0,916, 0,999)	1,054 (0,939, 1,184)	0,970 (0,932, 1,008)
Temperatura neta efectiva (°C) [<5,02 °C]			
5,02 °C – 12,46 °C	1,045 (1,004, 1,087)		1,013 (0,976, 1,052)
12,46 °C – 19,97 °C	1,253 (1,142, 1,373)	1,852 (1,409, 2,434)	1,284 (1,185, 1,391)
> 19,97 °C	0,620 (0,129, 2,112)	1,696 (1,271, 2,261)	1,277 (1,150, 1,418)
Pressió atmosfèrica (hPa) [<904,15 hPa]			
904,15 hPa – 955,64 hPa	0,971 (0,929, 1,016)	1,222 (0,953, 1,579)	0,966 (0,927, 1,006)
955,64 hPa – 993,98 hPa	1,006 (0,946, 1,070)	1,221 (0,948, 1,584)	0,976 (0,925, 1,030)
> 993,98	1,003 (0,936, 1,074)	1,381 (1,058, 1,813)	0,977 (0,919, 1,038)
Irradiació solar global (W/m²) [<101,54 W/m²]			
101,54 W/m² – 171,40 W/m²	0,949 (0,922, 0,978)	0,944 (0,853, 1,045)	0,956 (0,930, 0,983)
171,40 W/m² – 294,51 W/m²	0,870 (0,834, 0,907)	1,099 (0,983, 1,228)	0,913 (0,879, 0,949)
> 294,51 W/m²	0,987 (0,909, 1,070)	0,950 (0,820, 1,100)	0,961 (0,903, 1,023)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat .

En tots els casos, promig dels retards 1 a 10 de la variable.

Taula 8. Associació entre contaminants atmosfèrics i els casos hospitalitzats.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
PM10 (μm^3) [$<20,80 \mu\text{m}^3$]			
20,80 μm^3 – 23,34 μm^3	0,988 (0,873, 1,117)	1,128 (0,905, 1,407)	0,968 (0,862, 1,087)
23,34 μm^3 – 25,89 μm^3	0,933 (0,812, 1,071)	0,880 (0,693, 1,118)	0,957 (0,840, 1,089)
> 25,89 μm^3	0,952 (0,811, 1,118)	0,940 (0,695, 1,271)	0,940 (0,808, 1,093)
NO2 (μm^3) [$<21,93 \mu\text{m}^3$]			
21,93 μm^3 - 25,77 μm^3	1,203 (1,059, 1,366)	1,244 (1,002, 1,545)	1,180 (1,047, 1,330)
25,77 μm^3 – 30,65 μm^3	1,040 (0,906, 1,195)	0,937 (0,737, 1,190)	1,026 (0,901, 1,168)
> 30,64 μm^3	1,145 (0,963, 1,362)	1,373 (0,925, 1,821)	1,140 (0,969, 1,342)
O3 (μm^3) [$<45,98 \mu\text{m}^3$]			
45,98 μm^3 – 52,96 μm^3	0,933 (0,817, 1,065)	0,894 (0,712, 1,120)	0,936 (0,826, 1,059)
52,96 μm^3 – 60,37 μm^3	0,900 (0,786, 1,030)	1,319 (1,042, 1,672)	1,130 (1,001, 1,284)
> 60,37 μm^3	1,020 (0,882, 1,179)	0,921 (0,715, 1,186)	1,015 (0,885, 1,164)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat.

Ajustat per mobilitat (percentatge de persones que es mouen en l'ABS; i percentatge de persones que es mouen fora de la mateixa àrea, respecte a les que es mouen en l'ABS) (promig dels retards 1 a 10 i quartils); renda mitjana per persona en l'ABS (quartils); taxa d'atur en l'ABS (quartils); densitat de la població en l'ABS (quartils); percentatge d'infrahabitatges en l'ABS (quartils); percentatge de persones de 65 anys o més en l'ABS; percentatge d'estrangers de països de baixa renda (quartils); i confinament (en el model per a tot el període).

Taula 9. Associació entre les variables meteorològiques i els ingressos en UCI.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
Temperatura (°C) [<8,61 °C]			
8,61 °C – 15,58 °C	1,146 (0,994, 1,322)		1,101 (0,960, 1,264)
15,58 °C – 23,50 °C	1,595 (1,138, 2,226)	2,720 (0,802, 9,712)	1,607 (1,191, 2,167)
> 23,50 °C	2,518 (2e-17, 2e+07)	2,048 (0,562, 7,829)	1,660 (1,083, 2,546)
Humitat relativa (%) [<61,82%]			
61,82% - 70,65%	0,948 (0,808, 1,115)	0,766 (0,551, 1,058)	0,826 (0,716, 0,954)
70,65% - 79,37%	0,988 (0,838, 1,168)	0,729 (0,465, 1,125)	0,855 (0,736, 0,995)
> 79,37%	1,108 (0,925, 1,330)	1,245 (0,503, 2,759)	0,954 (0,808, 1,129)
Velocitat del vent a 2 m (m/s) [<0,56 m/s]			
0,56 m/s – 0,76 m/s	1,030 (0,932, 1,139)	0,804 (0,525, 1,238)	1,010 (0,916, 1,114)
0,76 m/s – 1,08 m/s	0,997 (0,882, 1,127)	1,046 (0,693, 1,598)	1,002 (0,891, 1,125)
> 1,08 m/s	0,967 (0,846, 1,105)	1,256 (0,782, 2,031)	0,988 (0,870, 1,123)
Temperatura neta efectiva (°C) [<5,02 °C]			
5,02 °C – 12,46 °C	1,093 (0,961, 1,243)		1,067 (0,941, 1,209)
12,46 °C – 19,97 °C	1,740 (1,244, 2,424)	3,083 (0,899, 1,18e+01)	1,756 (1,297, 2,376)
> 19,97 °C	7,105 (0,682, 4,25e+01)	2,879 (0,792, 1,16e+01)	1,967 (1,292, 2,998)
Pressió atmosfèrica (hPa) [<904,15 hPa]			
904,15 hPa – 955,64 hPa	0,880 (0,771, 1,004)	0,834 (0,346, 2,235)	0,865 (0,760, 0,983)
955,64 hPa – 993,98 hPa	0,817 (0,683, 0,976)	0,861 (0,353, 2,332)	0,822 (0,694, 0,973)
> 993,98	0,777 (0,640, 0,944)	0,856 (0,332, 2,425)	0,772 (0,640, 0,930)
Irradiació solar global (W/m²) [<101,54 W/m²]			
101,54 W/m² – 171,40 W/m²	0,965 (0,879, 1,059)	0,944 (0,586, 1,536)	0,960 (0,876, 1,052)
171,40 W/m² – 294,51 W/m²	0,933 (0,822, 1,058)	1,120 (0,700, 1,834)	0,952 (0,843, 1,073)
> 294,51 W/m²	1,011 (0,746, 1,360)	0,914 (0,518, 1,632)	0,993 (0,792, 1,2410)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat.

En tots els casos, promig dels retards 1 a 10 de la variable.

Taula 10. Associació entre contaminants atmosfèrics i els ingressos a l'UCI.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
PM₁₀ (μ/m³) [<20,80 μ/m³]			
20,80 μ/m³ – 23,34 μ/m³	0,895 (0,764, 1,049)	0,979 (0,573, 1,671)	0,904 (0,773, 1,057)
23,34 μ/m³ – 25,89 μ/m³	0,905 (0,759, 1,079)	1,018 (0,554, 1,858)	0,924 (0,776, 1,100)
> 25,89 μ/m³	0,911 (0,746, 1,113)	0,814 (0,411, 1,600)	0,900 (0,739, 1,098)
NO₂ (μ/m³) [<21,93 μ/m³]			
21,93 μ/m³ - 25,77 μ/m³	1,208 (1,027, 1,422)	1,206 (0,692, 2,126)	1,208 (1,028, 1,420)
25,77 μ/m³ – 30,65 μ/m³	1,241 (1,038, 1,485)	1,228 (0,671, 2,265)	1,254 (1,051, 1,498)
> 30,64 μ/m³	1,318 (1,055, 1,646)	1,695 (0,802, 3,636)	1,361 (1,093, 1,696)
O₃ (μ/m³) [<45,98 μ/m³]			
45,98 μ/m³– 52,96 μ/m³	0,965 (0,821, 1,134)	0,835 (0,482, 1,448)	0,950 (0,810, 1,115)
52,96 μ/m³ – 60,37 μ/m³	1,005 (0,850, 1,189)	0,944 (0,534, 1,658)	0,993 (0,841, 1,172)
> 60,37 μ/m³	1,032 (0,858, 1,242)	1,491 (0,800, 2,788)	1,054 (0,878, 1,265)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat.

Ajustat per mobilitat (percentatge de persones que es mouen en l'ABS; i percentatge de persones que es mouen fora de la mateixa àrea, respecte a les que es mouen en l'ABS) (promig dels retards 1 a 10 i quartils); renda mitjana per persona en l'ABS (quartils); taxa d'atur en l'ABS (quartils); densitat de la població en l'ABS (quartils); percentatge d'infrahabitatges en l'ABS (quartils); percentatge de persones de 65 anys o més en l'ABS; percentatge d'estrangers de països de baixa renda (quartils); i confinament (en el model per a tot el període).

Taula 11. Associació entre les variables meteorològiques i els morts.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
Temperatura (°C) [<8,61 °C]			
8,61 °C – 15,58 °C	0,914 (0,835, 1,002)		0,922 (0,845, 1,007)
15,58 °C – 23,50 °C	1,204 (0,964, 1,498)	1,343 (1,578, 7,506)	1,263 (1,0354, 1,539)
> 23,50 °C	0,693 (0,670, 4,12e+01)	2,921 (1,274, 6,731)	1,356 (1,021, 1,801)
Humitat relativa (%) [<61,82%]			
61,82% - 70,65%	1,091 (0,972, 1,227)	0,842 (0,666, 1,063)	0,883 (0,799, 0,978)
70,65% - 79,37%	1,138 (1,009, 1,286)	0,920 (0,659, 1,278)	0,932 (0,838, 1,037)
> 79,37%	1,144 (1,004, 1,306)	0,771 (0,375, 1,497)	0,941 (0,837, 1,060)
Velocitat del vent a 2 m (m/s) [<0,56 m/s]			
0,56 m/s – 0,76 m/s	0,900 (0,845, 0,958)	0,718 (0,534, 0,967)	0,890 (0,837, 0,946)
0,76 m/s – 1,08 m/s	0,923 (0,853, 0,999)	0,837 (0,619, 1,136)	0,926 (0,859, 0,998)
> 1,08 m/s	0,926 (0,846, 1,013)	1,081 (0,766, 1,532)	0,953 (0,875, 1,037)
Temperatura neta efectiva (°C) [<5,02 °C]			
5,02 °C – 12,46 °C	1,076 (0,991, 1,168)		1,067 (0,985, 1,155)
12,46 °C – 19,97 °C	1,757 (1,407, 2,188)	3,767 (1,744, 8,291)	1,773 (1,448, 2,168)
> 19,97 °C	0,807 (0,418, 1,196)	3,220 (1,430, 7,376)	1,871 (1,415, 2,476)
Pressió atmosfèrica (hPa) [<904,15 hPa]			
904,15 hPa – 955,64 hPa	0,860 (0,781, 0,946)	0,713 (0,345, 1,577)	0,857 (0,782, 0,937)
955,64 hPa – 993,98 hPa	1,037 (0,913, 1,177)	0,718 (0,344, 1,600)	0,961 (0,855, 1,081)
> 993,98	1,067 (0,927, 1,228)	0,995 (0,460, 2,289)	0,994 (0,871, 1,134)
Irradiació solar global (W/m²) [<101,54 W/m²]			
101,54 W/m² – 171,40 W/m²	0,852 (0,801, 0,906)	1,853 (1,327, 2,627)	0,882 (0,831, 0,936)
171,40 W/m² – 294,51 W/m²	0,718 (0,656, 0,785)	1,598 (1,096, 2,355)	0,760 (0,698, 0,827)
> 294,51 W/m²	0,703 (0,578, 0,851)	1,008 (0,635, 1,610)	0,693 (0,594, 0,808)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat.

En tots els casos, promig dels retards 1 a 10 de la variable

Taula 12. Associació entre contaminants atmosfèrics i els morts.

	1era onada 28 febrer-30 juny	Període vall 1 juliol – 31 agost	Tot el període 28 febrer – 31 agost
PM₁₀ (μ/m³) [<20,80 μ/m³]			
20,80 μ/m³ – 23,34 μ/m³	0,951 (0,786, 1,152)	0,869 (0,502, 1,495)	0,950 (0,794, 1,137)
23,34 μ/m³ – 25,89 μ/m³	0,962 (0,777, 1,191)	0,665 (0,363, 1,207)	0,972 (0,796, 1,188)
> 25,89 μ/m³	0,950 (0,742, 1,216)	0,835 (0,388, 1,787)	0,913 (0,724, 1,152)
NO₂ (μ/m³) [<21,93 μ/m³]			
21,93 μ/m³ – 25,77 μ/m³	1,090 (0,895, 1,327)	1,204 (0,701, 2,079)	1,068 (0,887, 1,285)
25,77 μ/m³ – 30,65 μ/m³	1,076 (0,868, 1,334)	1,096 (0,598, 2,013)	1,037 (0,848, 1,270)
> 30,64 μ/m³	1,094 (0,837, 1,430)	0,672 (0,326, 1,382)	1,091 (0,848, 1,403)
O₃ (μ/m³) [<45,98 μ/m³]			
45,98 μ/m³ – 52,96 μ/m³	0,796 (0,550, 1,042)	1,108 (0,625, 1,966)	0,809 (0,568, 1,050)
52,96 μ/m³ – 60,37 μ/m³	0,822 (0,669, 1,011)	0,895 (0,481, 1,650)	0,839 (0,690, 1,019)
> 60,37 μ/m³	0,943 (0,754, 1,179)	1,238 (0,649, 2,354)	0,959 (0,777, 1,184)

El interval de credibilitat al 95% no conté la unitat, El interval de credibilitat al 90% no conté la unitat.

Ajustat per mobilitat (percentatge de persones que es mouen en l'ABS; i percentatge de persones que es mouen fora de la mateixa àrea, respecte a les que es mouen en l'ABS) (promig dels retards 1 a 10 i quartils); renda mitjana per persona en l'ABS (quartils); taxa d'atur en l'ABS (quartils); densitat de la població en l'ABS (quartils); percentatge d'infrahabitatges en l'ABS (quartils); percentatge de persones de 65 anys o més en l'ABS; percentatge d'estrangers de països de baixa renda (quartils); i confinament (en el model per a tot el període).

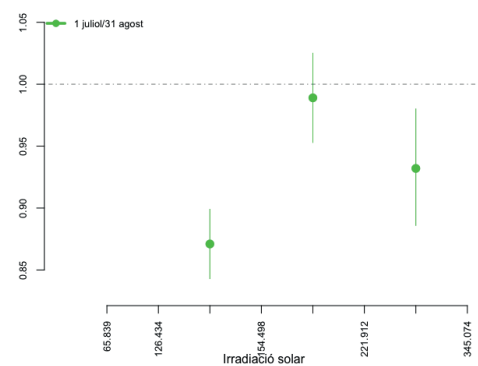
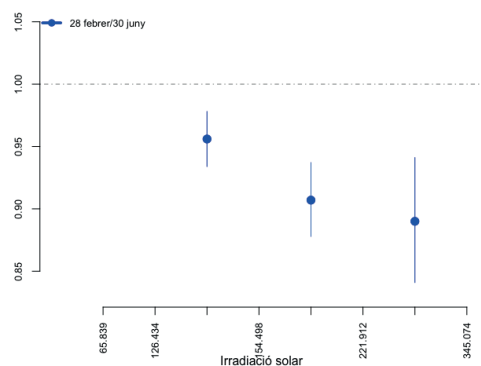
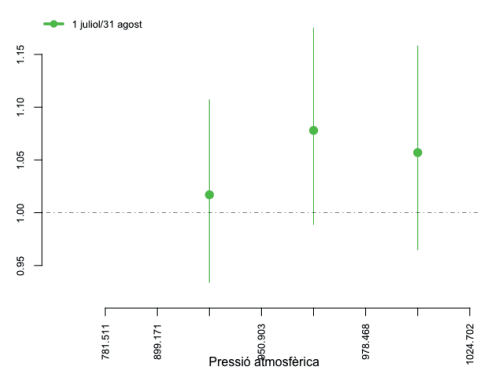
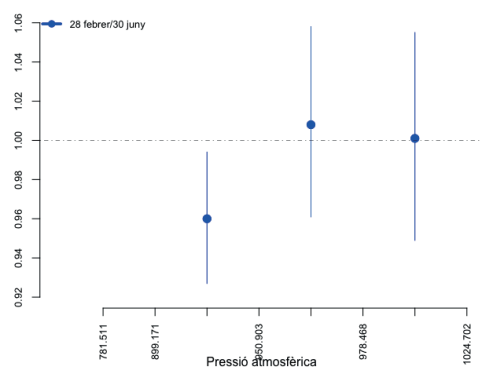
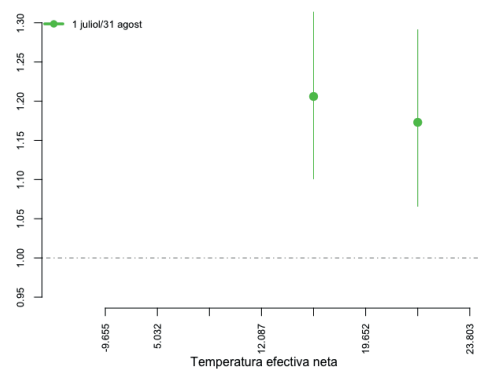
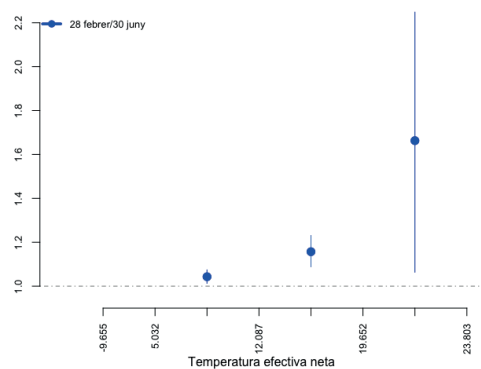


Figura 45. Associació entre les variables meteorològiques i els casos positius per PCR.

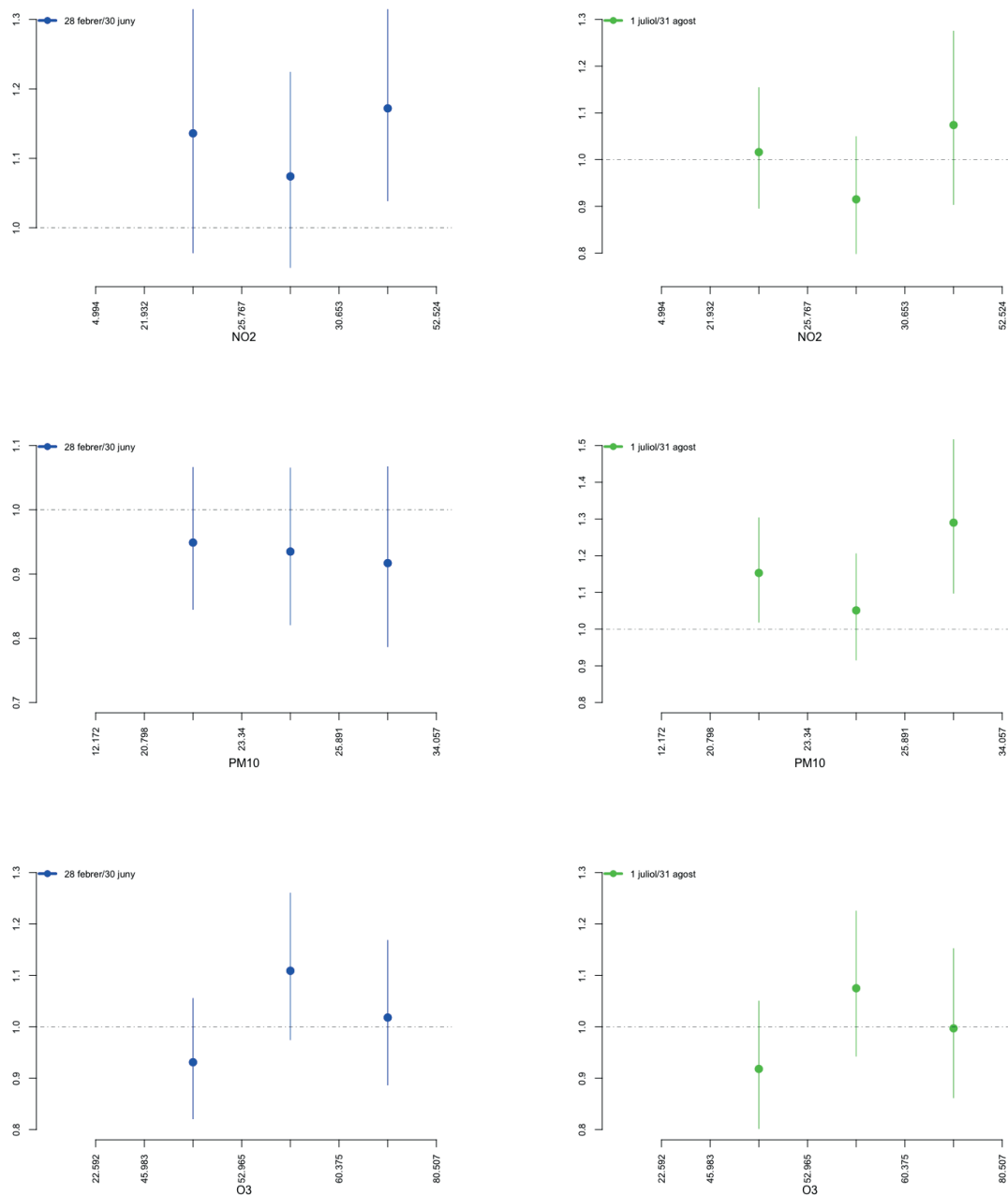


Figura 46. Associació entre els contaminants atmosfèrics i els casos positius per PCR.

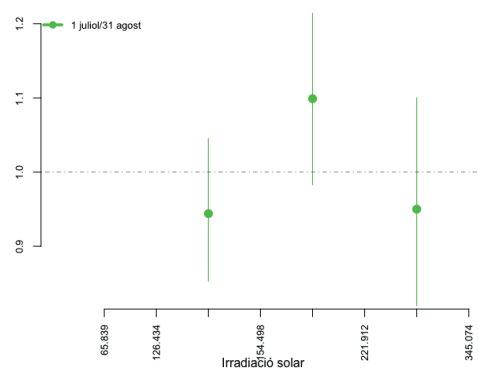
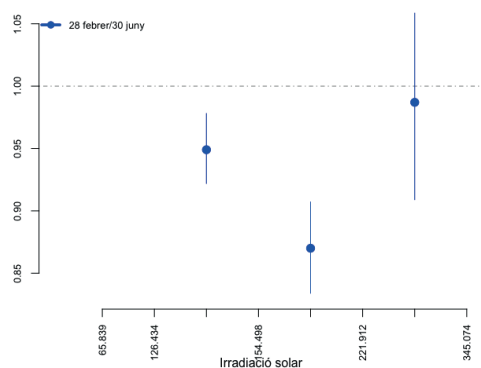
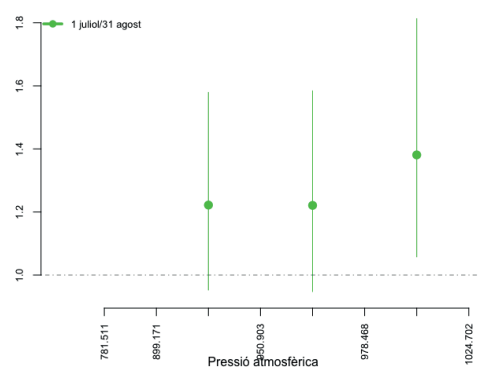
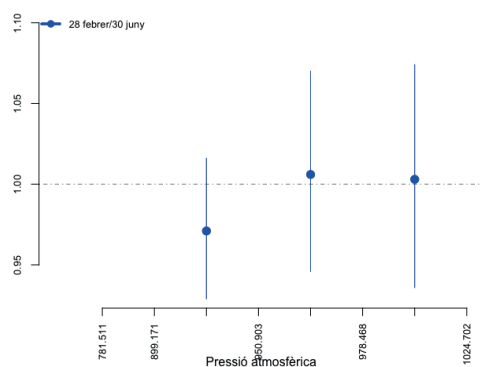
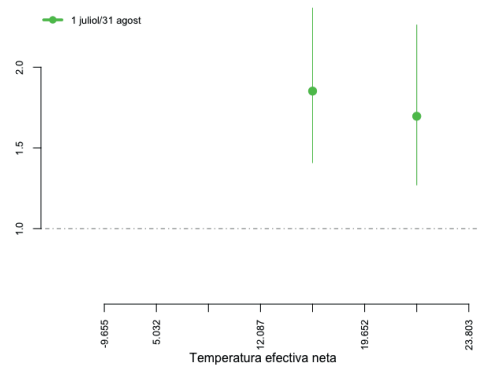
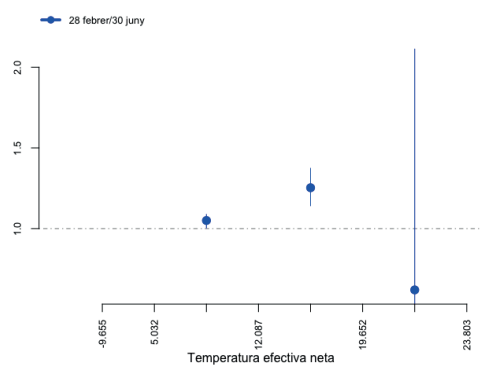


Figura 47. Associació entre les variables meteorològiques i els casos hospitalitzats.

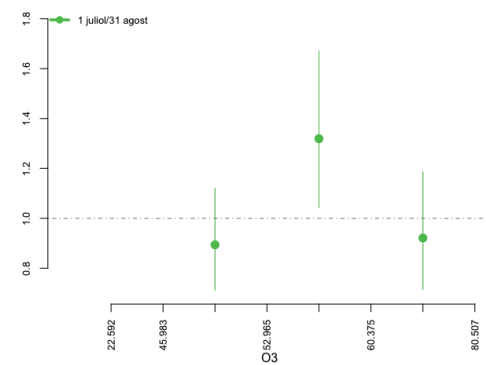
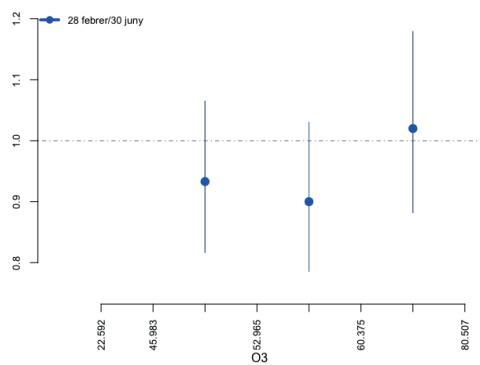
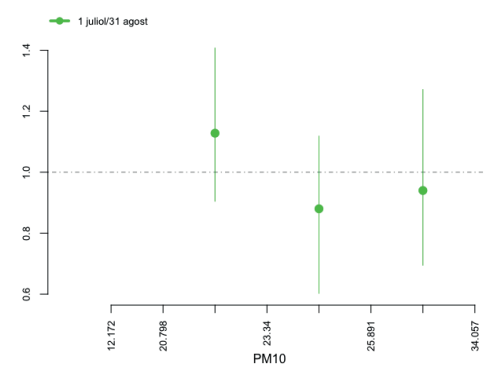
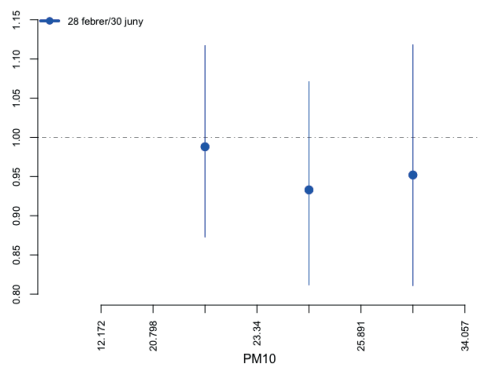
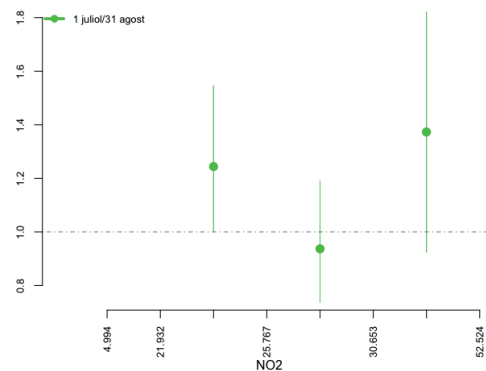
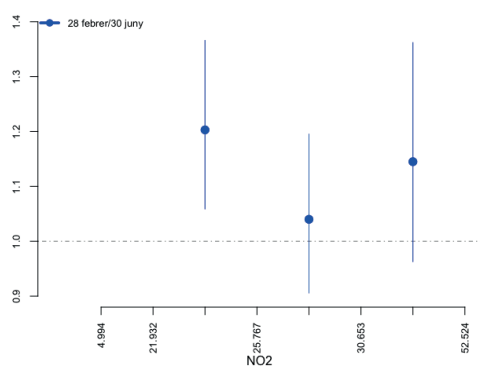


Figura 48. Associació entre els contaminants atmosfèrics i els casos hospitalitzats.

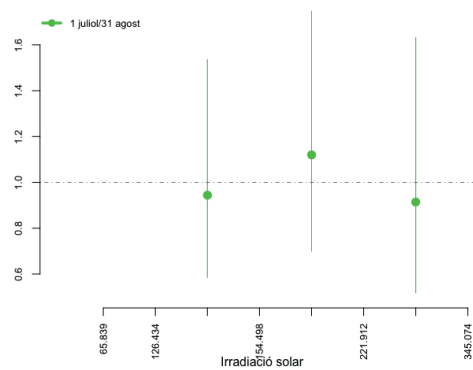
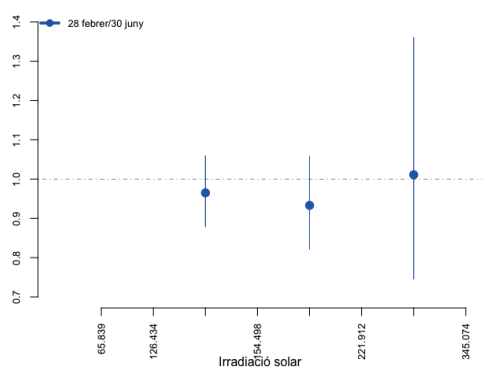
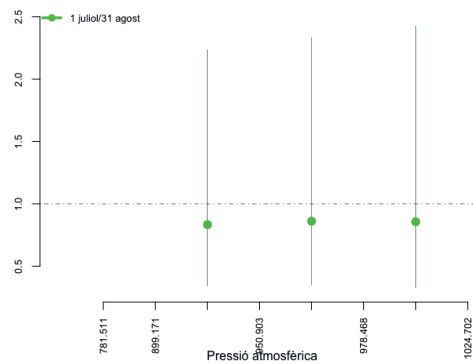
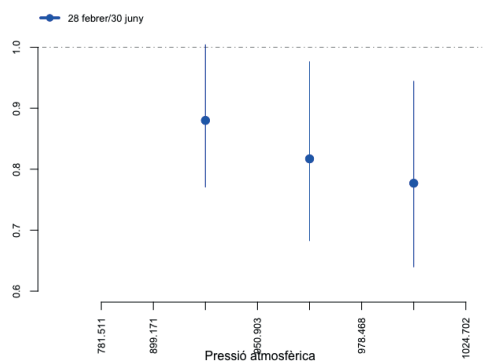
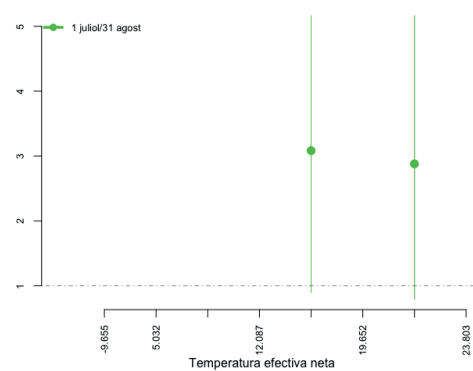
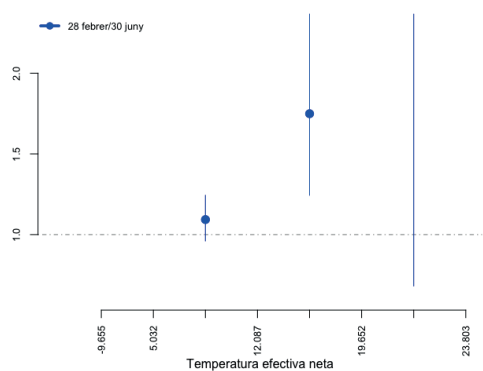


Figura 49. Associació entre les variables meteorològiques i els ingressos UCI.

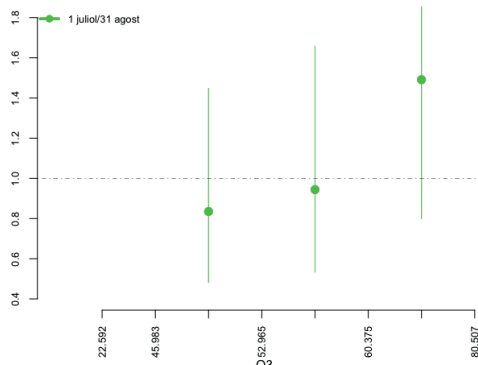
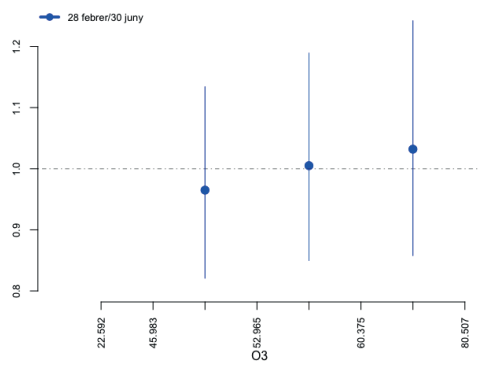
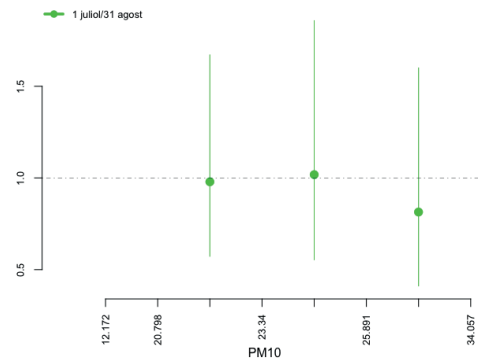
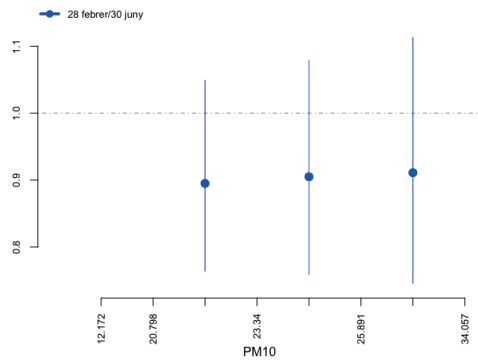
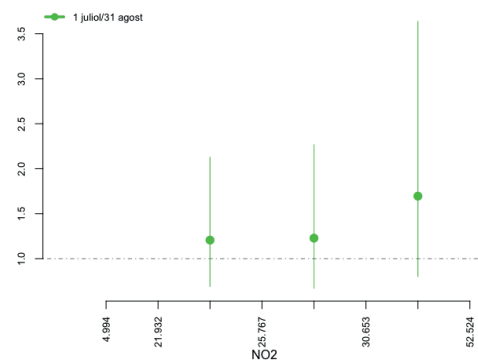
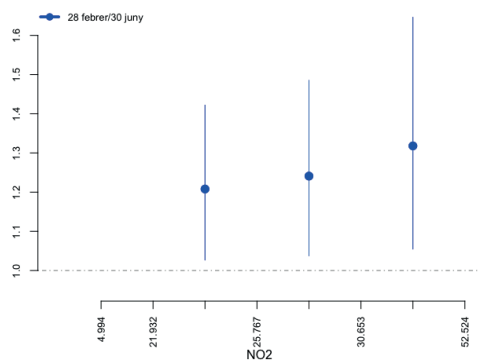


Figura 50. Associació entre els contaminants atmosfèrics i els ingressos UCI.

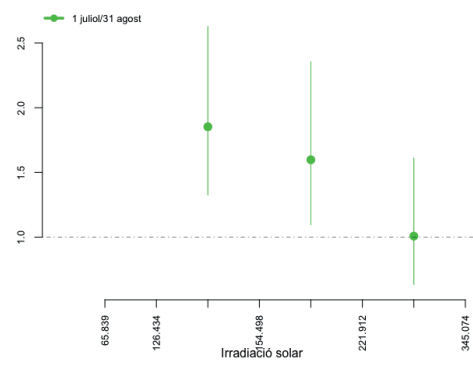
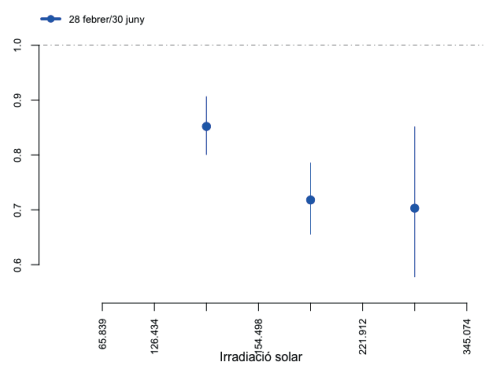
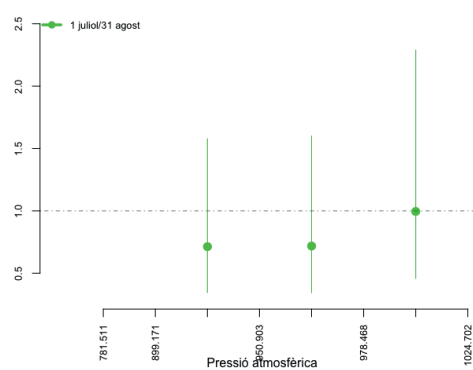
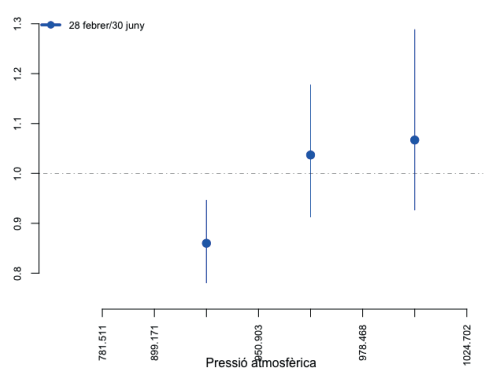
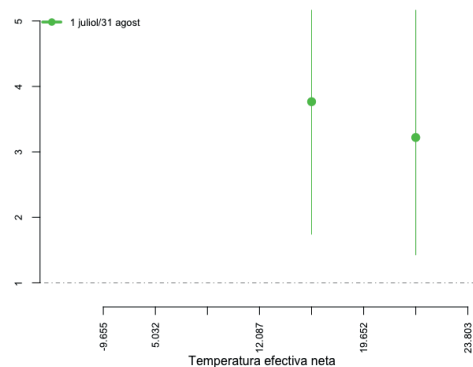
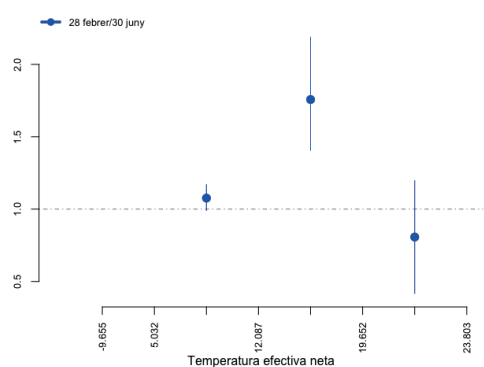


Figura 51. Associació entre les variables meteorològiques i els morts.

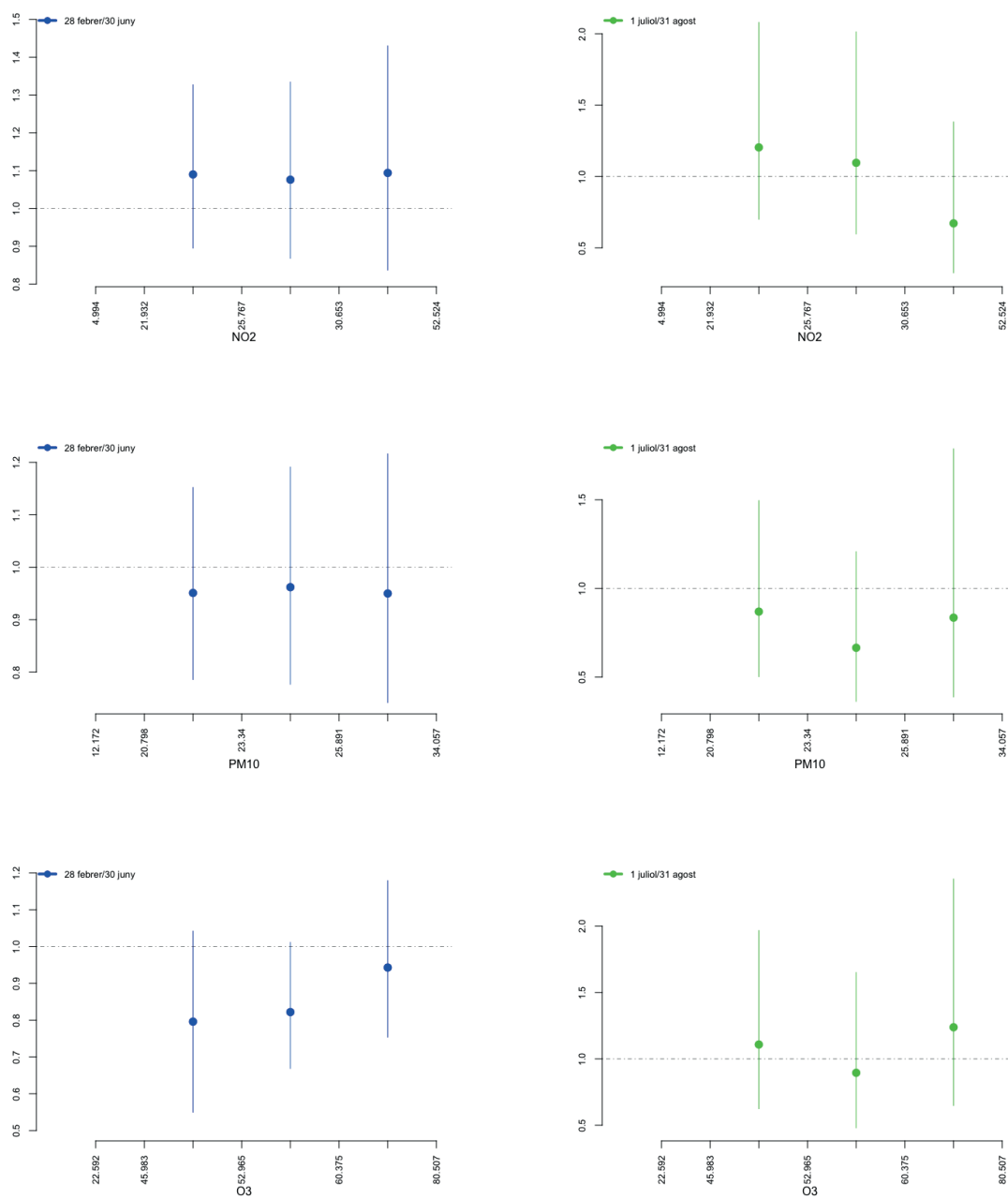


Figura 52. Associació entre els contaminants atmosfèrics i els morts.

Annex III: Model Epidemiològic amb Equacions Diferencials en detall

1 COVID-19 and input data

On 31 December 2019, a cluster of cases of pneumonia of unknown cause, in the city of Wuhan, Hubei province in China, was reported to the World Health Organisation. In January 2020, a previously unknown new virus was identified [1], subsequently named the 2019 novel coronavirus, and samples obtained from cases and analysis of the virus' genetics indicated that this was the cause of the outbreak. This novel coronavirus was named Coronavirus Disease 2019 (COVID-19) by WHO in February 2020 [2]. The virus is referred to as SARSCoV-2 and the associated disease is COVID-19 [3].

In January 2020 the first case of COVID-19 was reported in Spain [4]. As of today, 19th of January of 2021, there have been 2.370.742 diagnosed cases and 54.173 people deceased in Spain since the beginning of 2020.

In general, at the beginning of epidemics due to emerging diseases, the system detects only the most serious cases, which are those that go to hospitals and can be diagnosed and notified to the surveillance network. For this reason, it is common to observe at the beginning "more serious cases and greater lethality", since mild cases are not detected and therefore are not part of the denominator.

There are some peculiarities about the dynamics of COVID-19 that we would like to mention:

- **Profile ICU admission.** People older than 50 years old have a higher presence in ICU than younger people⁴. However, according to the data reported on the 7th of May of 2020 [5], patients admitted at ICU were significantly younger than those hospitalized without ICU admission (mean age 65 vs. 70), being the percentage of patients older than 80 years old admitted at ICU of 5% versus 28% in the group of hospitalized patients without ICU admission.
- **Incidence of COVID-19 in health or social health personnel.** 4.7% of the cases are health or social health personnel, this percentage being significantly higher among women than among men (7% vs 2.2%). 78% of health or social health personnel with COVID-19, notified since May 11, are women⁵.

Bi et al. showed that children are at similar risk of infection as the general population, though less likely to have severe symptoms; hence should be considered in analyses of transmission and control [6].

⁴ More information here: <https://bit.ly/39RFV8T>

⁵ More information here: <https://bit.ly/39RFV8T>

1.1 Disease dynamics

To understand the spreading dynamics of an infectious disease, there are three relevant time periods that should be carefully distinguished:

- **Latent or Pre-infectious period.** It is the time interval between when an individual is infected by a pathogen and when the individual becomes infectious, i.e. capable of transmitting pathogens to other susceptible individuals.
- **Infectious period.** It is the time interval during which the host is infectious, i.e. the pathogens can be transmitted directly or indirectly from the infected host to another individual, defined as the period from the end of the latent period until the time when the host can no longer transmit the infection to other individuals.

- **Incubation period.** It is the time elapsed between exposure to a pathogen and when symptoms and signs are first apparent.

For COVID-19, the symptoms of the disease appear after the individual becomes infectious. In this scenario, the latent period has a shorter duration than the incubation period, which means that the individuals can infect others for some time without showing any noticeable symptoms. This early or mild stage of infection whose symptoms stay below the level of clinical detection is called subclinical infection and the individual concerned is called an asymptomatic carrier of the disease. Figure 1 displays the infectiousness level of an asymptomatic carrier and a severe symptomatic patient.

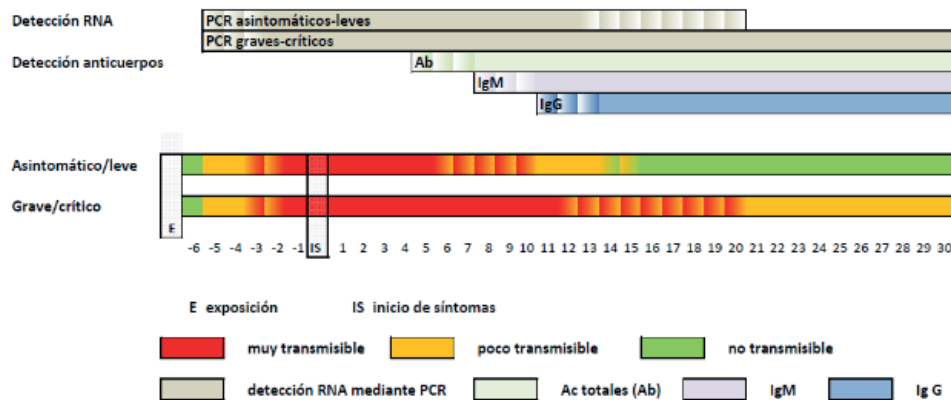


Figure 1. Average periods of transmissibility according to the severity of COVID-19 cases and detection periods of SARS-CoV-2 RNA by PCR and of antibodies by serological techniques. Source: [7].

The serial interval refers to the time between successive cases in a chain of transmission. The mean serial interval in numerous epidemiological observations has been shorter than the incubation period (Figure 2). On the basis of these observations and the cases detected in the exhaustive contact studies [8, 9, 10, 11], it was initially known that the transmission of the infection began 1-2 days before the onset of symptoms [12].

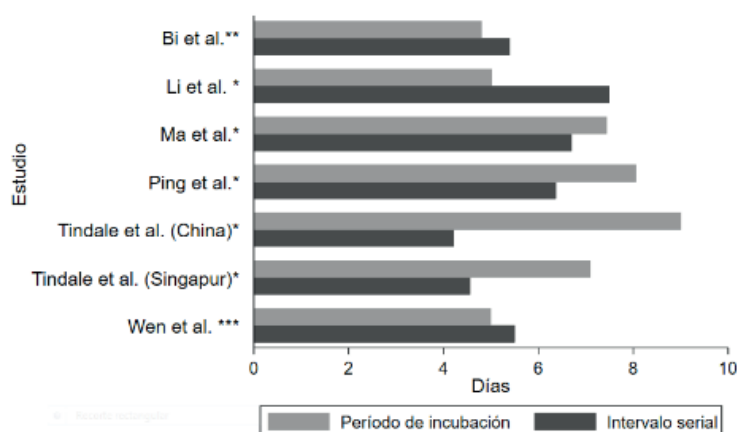


Figure 2. Comparison of the incubation period with the serial interval in studies that estimated both values.

* Incubation period and serial interval mean values;

** Median incubation period and average serial interval;

*** Incubation period and serial interval mean values. Source: [7].

Given that, symptomatic surveillance-based disease control measures (such as isolation, contact tracing, enhanced hygiene, etc.) are likely to have their effectiveness reduced, because a significant portion of the transmission may take place before the onset of symptoms and this have to be taken into account when designing control measures [13, 14].

Immunity

To assess immunity for COVID-19 we should look at the presence of AntiSARS-CoV-2 immunoglobulin G (IgG) in our body fluids. IgG is a type of antibody that provides the majority of antibody-based immunity against invading pathogens, meaning that it is the main potentiator of humoral immunity. The presence of IgG indicates a previous infection. In other words, the detection of a positive level of Anti-SARS-CoV-2 IgG in a person will imply the previous infection of COVID-19 and the current immunity to the virus. Some people — because they had mild or no symptoms, for example — might have developed antibodies that are too weak to prevent re-infection. Conversely, others who have low levels of IgG may still be protected.

At the moment, there are no resounding scientific studies that ensure that every infected person develops an effective cellular immune response, in such a way that it generates antibodies that neutralize the virus and eradicate it from the body, and with memory acquisition [15].

Most persons infected with SARS-CoV-2 display an antibody response between day 10 and day 21 after infection. Detection in mild cases can take longer time (four weeks or more) and in a small number of cases antibodies (i.e., IgM, IgG) are not detected at all (at least during the studies' time scale). Based on the currently available data, the IgM and IgG antibodies to SARS-CoV-2 develop between 6–15 days post disease onset [16, 17, 18, 19, 20, 21]. The median seroconversion time for total antibodies, IgM and then IgG were day-11, day-12 and day-14 post symptom onset, respectively. The presence of antibodies was detected in <40% among patients within 1 week from onset, and rapidly increased to 100% (total antibodies), 94.3% (IgM) and 79.8% (IgG) from day-15 after onset [17].

The longevity of the antibody response is still unknown, but it is known that antibodies to other coronaviruses wane over time (range: 12 – 100 weeks from the onset of symptoms; for the severe acute respiratory syndrome (SARS), immunity lasted for 2 years on average [22]) and

homologous re-infections have been shown [23]. SARS-CoV-2 IgM and IgG antibody levels may remain over the course of seven weeks [24] or at least in 80% of the cases until day 49 [25]. In comparison, 90% and 50% of SARS-CoV-1 infected patients have been shown to maintain IgG antibodies for two and three years respectively [22]. In addition, it could be important to detect nasal IgA antibodies, as the serum IgA antibodies were not raised, but IgA persisted in the nasal mucosa one year post-infection for seasonal coronavirus 229E [26].

To bring this pandemic to an end, a large share of the world needs to be immune to the virus. The safest way to achieve this is with a vaccine. Within less than 12 months after the beginning of the COVID-19 pandemic, several research teams rose to the challenge and developed vaccines that protect from SARS-CoV-2, the virus that causes COVID-19. Now the challenge is to make these vaccines available to people around the world. It will be key that people in all countries — not just in rich countries — receive the required protection.

1.2 Mobility dynamics

Iacus et al. [27] demonstrated that human mobility, derived by mobile data, is highly correlated with the spread of COVID-19 in the initial phase of the outbreak. In the case study of France, we have found that mobility can explain from 52% up to 92% of the excess deaths reasonably linked to the COVID-19 outbreak.

1.3 Input data

The data used to build the epidemic model on will depend on the population of study:

1. Catalonia.

The *Agència de Qualitat i Avaluació Sanitàries de Catalunya* (AQUAS)⁶ is the responsible partner of channeling all the data from the distinct organizations that form the healthcare system in Catalonia. AQUAS provide a dataset that contains COVID-19 incidence per *Àrea Bàsica de Salut* (ABS), which could be mapped geographically. In particular, we have the following data per ABS:

- Age group: [0, 25), [25, 50), and ≤ 50 years
- Date of first positive PCR
- Number of positive PCRs
- Average days at ICU, excluding hospitalization
- In-hospital mortality
- Out-hospital mortality in less than 30 days
- Out-hospital mortality from 30 to 60 days
- Out-hospital mortality in more than 60 days

And for each ABS, age group, and date of first positive PCR, we have the following information:

- Number of hospitalizations: if one person is hospitalized twice, two hospitalizations will be counted
- Average days of hospitalizations
- Number of ICU admissions

2. Spain.

The *Instituto Nacional de Estadística* (INE)^[2] is the party that centralizes the data about COVID-19 incidence of Spain. The available data is based on Spain provinces' and it gives information about:

- Number of COVID-19 cases
- Number of positive PCR
- Hospitalizations
- ICU admissions
- Mortality

In relation to the mobility model, the data from Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA)⁷ is used to depict the mobility reality of Catalonia, while data from INE [28] is used to model the mobility of Spain.

2 Mobility model

The population is distributed in a set of patches (meta-population). Individuals are segregated into N_g different age-groups. Let N be the number of patches in the region under study and the initial population in each patch for $i \in \{1, \dots, N\}$ of age-range g for $g \in \{1, \dots, N_g\}$, so that the total number of individuals in the system is :

$$M = \sum_{i=1}^N \sum_{g=1}^{N_g} n_{i0}^g$$

The number of individuals in each patch may be different.

Let h be the time-step in hours so that $N_h = 24/H$ is the total number of time-slots defined in each day. Mobility of individuals might change according to the time of day. Thus, mobility is defined by time-slots s , where $s \in \{1, \dots, N_h\}$. In particular, we consider three kind of movements:

- move from one patch to another, depicted by $p_i^{g,s}$, which is the fraction of individuals of age-group g moving from patch i to other patches;
- move within the patch, depicted by $r_i^{g,s}$; and
- stay at home, depicted by $q_i^{g,s}$;

so that: $p_i^{g,s} + q_i^{g,s} + r_i^{g,s} = 1$.

The mobility model defines the mobility patters between individuals. It could be defined as an original Movement-Interaction-Return (MIR) model or an extended MIR model. In the MIR model, individuals belonging to patch i move to other patches and return to patch i after a time step. Thus, each patch is always composed of the same individuals. Instead, in the extended MIR model, individuals are allowed to start outside their home-patch. Consequently, each patch at each time step is (potentially) composed of individuals from different patches. Individuals in a patch different from their home-patch, either return back home or choose to stay in the patch with a certain probability [29].

For individuals of age-group $g \in \{1, \dots, N_g\}$ in patch $i \in \{1, \dots, N\}$ and current time-slot $s \in \{1, \dots, N_h\}$, the mobility model will define:

- Mobility probabilities. The probability of an individual of age-group g resident in patch i moving from patch j to patch k . If $i = j$, then this is defined by a usual Origin-Destination matrix.

- Displacement rates. The rate of individuals of age-group g from patch i moving between patches at time-slot s of the day, $p_i^{g,s}$. The rate of those individuals staying at their household, $q_i^{g,s}$, or moving within their patch with a rate of $r_i^{g,s}$, so that: $p_i^{g,s} + q_i^{g,s} + r_i^{g,s} = 1$. For example, during the night slot it could be assumed that a huge fraction of the population stay at their household. Thus, contacts are only between individuals within the same household.
- Contacts matrix. Average number of contacts (defined as contacts closer than 1.5 m) between individuals of each age-group, depicted as C_{gf} for all $g, f \in \{1, \dots, N_g\}$

Mobility restrictions and social distancing rules significantly affect the average number of contacts between individuals as well as mobility and displacement probabilities.

3 Epidemic model

3.1 Compartmental epidemic model

The epidemic model proposed here has eight epidemiological compartments, as shown in Figure 3 for each of the patches composing the meta-population: Susceptible (S), Exposed (E), Asymptomatic (A), Symptomatic (I), Hospitalized (H), Hospitalized in ICU (U), Recovered (R), and Dead (D)

(D). Both asymptomatic and symptomatic compartments together form the so-called infected compartment. In this study the traditional infected compartment have been segmented for the sake of an accurate characterization of the dynamics of COVID-19.

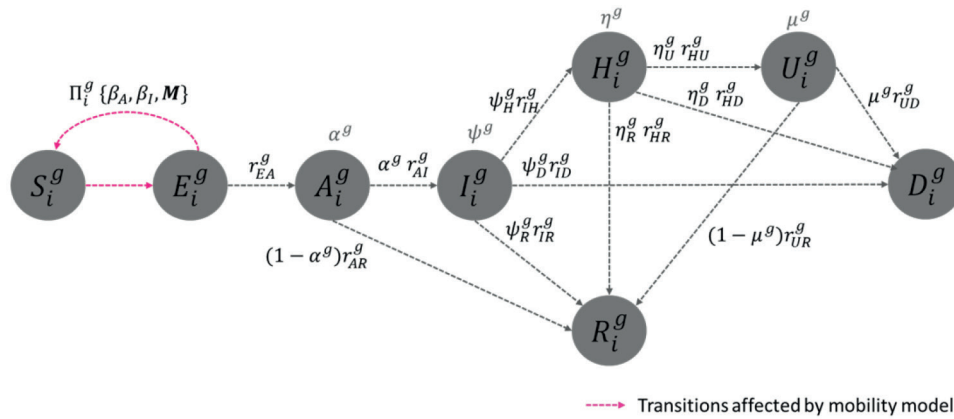


Figure 3. Compartmental epidemic model.
Susceptible (S), Exposed (E), Asymptomatic (A), Symptomatic (I), Hospitalized (H),
Hospitalized in ICU (U), Recovered (R), and Dead (D).

We consider the following continuous-time epidemic model describing the temporal evolution of individuals of age-group $g \in \{1, \dots, N_g\}$ in patch $i \in \{1, \dots, N\}$ and state $m \in \{S, E, A, I, H, U, D, R\}$ given by ordinary differential equations:

$$\frac{dS_i^g}{dt} = -\Pi_i^g(t) S_i^g(t) \quad (1)$$

$$\frac{dE_i^g}{dt} = \Pi_i^g(t) S_i^g(t) - r_{EA}^g E_i^g(t) \quad (2)$$

$$\frac{dA_i^g}{dt} = r_{EA}^g E_i^g(t) - \alpha^g r_{AI}^g A_i^g(t) - (1 - \alpha^g) r_{AR}^g A_i^g(t) \quad (3)$$

$$\frac{dI_i^g}{dt} = \alpha^g r_{AI}^g A_i^g(t) - \psi_H^g r_{IH}^g I_i^g(t) - \psi_D^g r_{ID}^g I_i^g(t) - \psi_R^g r_{IR}^g I_i^g(t) \quad (4)$$

$$\frac{dH_i^g}{dt} = \psi_H^g r_{IH}^g I_i^g(t) - \eta_U^g r_{HU}^g H_i^g(t) - \eta_D^g r_{HD}^g H_i^g(t) - \eta_R^g r_{HR}^g H_i^g(t) \quad (5)$$

$$\frac{dU_i^g}{dt} = \eta_U^g r_{HU}^g H_i^g(t) - \mu^g r_{UD}^g U_i^g(t) - (1 - \mu^g) r_{UR}^g U_i^g(t) \quad (6)$$

$$\begin{aligned} \frac{dR_i^g}{dt} = & (1 - \alpha^g) r_{AR}^g A_i^g(t) + \psi_R^g r_{IR}^g I_i^g(t) + \eta_R^g r_{HR}^g H_i^g(t) + \\ & + (1 - \mu^g) r_{UR}^g U_i^g(t) \end{aligned} \quad (7)$$

$$\frac{dD_i^g}{dt} = \psi_D^g r_{ID}^g I_i^g(t) + \eta_D^g r_{HD}^g H_i^g(t) + \mu^g r_{UD}^g U_i^g(t) \quad (8)$$

Where $M = \sum_{i=1}^N \sum_{g=1}^{N_g} (S_i^g + E_i^g + A_i^g + I_i^g + H_i^g + U_i^g + R_i^g + D_i^g)$ is the total population. Constant population size is assumed, including deaths. Note that for the analysis of rapid outbreaks that occur in very short periods of time, the population can be assumed to be in a constant state [30].

These equations correspond to a discrete-time dynamics, in which each timestep represents h hours and where the parameters are assumed to be nonnegative:

- r_{XY}^g is the rate of X individuals of age-group g turning into Y for $X, Y \in \{S, E, A, I, H, U, R, D\}$ per time-step, where r_{EI}^g is the latent period and r_{IH}^g is the incubation period;
- α^g is the probability that asymptomatic (A) individuals of age-group g develop symptoms (I);
- ψ_H^g is the probability that symptomatic (I) individuals of age-group g require hospitalization (H);
- ψ_D^g is the death (D) probability of symptomatic (I) individuals of agegroup g ;
- ψ_R^g is the probability that symptomatic (I) individuals of age-group g recover (R);
- η_U^g is the probability that hospitalized (H) individuals of age-group g require ICU (U);
- η_D^g is the death (D) probability of hospitalized (H) individuals of age-group g ;

- η_R^g is the probability that hospitalized (H) individuals of age-group g recover (R);
- μ^g is the death (D) probability of hospitalized individuals in ICU (U) of age-group g ;
- $\Pi_i^g(t)$ is the force of infection, i.e. the probability that a susceptible (S) individual of age-group g from patch i gets exposed (E) by contact with an infected individual, either asymptomatic (A) or symptomatic (I), within the meta-population at time t where t corresponds to a time slot $s \in \{1, \dots, N_g\}$ of the day [29, 31, 32, 33, 34]. This probability is defined as a function of:
 - Mobility probabilities from the mobility model, M ;
 - Displacement rates from the mobility model: $p_{i,s}^{g,s}, q_{i,s}^{g,s}, r_{i,s}^{g,s}$;
 - Contacts matrix, C_{gf} for all $g, f \in \{1, \dots, N_g\}$; o Number of individuals in patch i at time t , given by Equations 1-7 for all $g \in \{1, \dots, N_g\}$;
 - Number of infectious individuals, asymptomatic (A) or symptomatic (I), in patch i at time t , given by Equations 3 and 4 for all $g \in \{1, \dots, N_g\}$;
 - If homogeneous mixing is assumed between individuals, the transmission rate (infection rate) of asymptomatic (A) individuals, β_A , and symptomatic (I) individuals, β_I , should be considered. Instead, in contact based models, the probability of infection due to contact with an A individual, ρ_A , and an I individual, ρ_I , should be considered. Equation 9 gives the relation between these two parameters:
 $\beta = -C * \log(1 - \rho)$ (9)
 where C is the average contacts per unit time.
 - Average number of individuals in each household, infection rate within the household could be different to the one inside patches [35].

The derivation of the force of infection is explained in the following section.

3.2 Derivation of the force of infection

This section describes the derivation of the force of infection for single and meta-population models, including both homogeneous and contact based mixing assumptions.

3.2.1 Single population models

- A. Homogeneous mixing.** Assume a single population ($N = 1$) with homogeneous mixing between individuals. Then, the force of infection can be defined as:

$$\Pi(t) = \beta_A \sum_{f=1}^{N_g} \frac{A^f(t)}{n^f(t)} + \beta_I \sum_{f=1}^{N_g} \frac{I^f(t)}{n^f(t)} \quad (10)$$

where $n^i(t) = \sum X^i(t)$ where $X \in \{S, E, A, I, H, U, R\}$.

- B. Contact based mixing.** Assume a single population ($N = 1$) and let C_{gf} be the average number of contacts between individuals for $g, f \in \{1, \dots, N_g\}$. Then, the force of infection can be defined as a Bernoulli trial. Let x^g and y^g be the number of contacts between susceptible individuals of age-group g and asymptomatic and symptomatic individuals, respectively. Then, the force of infection can be defined as:

$$\Pi^g(t) = 1 - (1 - \rho_A)^{x^g} (1 - \rho_I)^{y^g} \quad (11)$$

The number of contacts with asymptomatic individuals is given by:

$$x^g = \sum_{f=1}^{N_g} C_{gf} \frac{A^f(t)}{n^f(t)} \quad (12)$$

Similarly, the number of contacts with infectious individuals is given by:

$$y^g = \sum_{f=1}^{N_g} C_{gf} \frac{I^f(t)}{n^f(t)}$$

Replacing expression 12 and 13 in Equation 11, the force of infection is given by:

$$\Pi^g(t) = 1 - \prod_{f=1}^{N_g} (1 - \rho_A)^{C_{gf} \frac{A^f(t)}{n^f(t)}} (1 - \rho_I)^{C_{gf} \frac{I^f(t)}{n^f(t)}} \quad (14)$$

3.2.2 Meta-population model with mobility

Assume a meta-population model with N patches. The probability that susceptible individuals of age-group g and patch i get exposed at time t is given by:

$$\Pi_i^g(t) = q_i^{g,s} P_{i(H)}^{g,s}(t) + r_i^{g,s} P_i^{g,s}(t) + p_i^{g,s} \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{M}_{ij} P_j^{g,s}(t) \quad (15)$$

where $P_i^{g,s}(t)$ denotes the probability that susceptible individuals of agegroup g get infected inside patch i , and $P_{i(H)}^{g,s}(t)$ denotes the probability of getting infected inside the household in patch i for $i \in \{0, \dots, N\}$. The three terms in Eq 15 refer, in order of appearance, to: the fraction of individuals getting infected inside their household at time t , the fraction of individuals that got infected inside their home-patch at time t , and the fraction of individuals that moved and got infected in their destination patch at time t .

- A. Homogeneous mixing.** In this case we assume homogeneous mixing between individuals that are currently in the same patch. Thus, the probability that a susceptible individual of patch i and age-group g gets infected is defined as:

$$P_i^{g,s}(t) = \beta_A \sum_{f=1}^{N_g} \frac{A_i^{f,s*}}{n_i^{f,s*}} + \beta_I \sum_{f=1}^{N_g} \frac{I_i^{f,s*}}{n_i^{f,s*}} \quad (16)$$

The former term of Equation 16 refers to the force of infection due to contact with asymptomatic individuals, while the latter refers to the force of infection due to contact with symptomatic individuals.

Similarly, the probability of infection within the household can be defined as:

$$P_{i(H)}^{g,s}(t) = \beta_A \frac{A_{i(H)}^{s*}}{n_{i(H)}^{s*}} + \beta_I \frac{I_{i(H)}^{s*}}{n_{i(H)}^{s*}} \quad (17)$$

B. Contact based mixing Here we assume a meta-population model with mobility and age-dependant contact patterns between individuals. We will first infer $P_i^{g,s}$ and later $P_{i(H)}^{g,s}$. Given x_i^g and y_i^g contacts between a susceptible individual of age-group g and asymptomatic and infected individuals in patch i , respectively, the probability of no infection can be defined as a Bernoulli trial:

$$\hat{P}_i^{g,s}(t) = (1 - \rho_A)^{x_i^{g,s}} (1 - \rho_I)^{y_i^{g,s}}.$$

Thus, the probability of infection of a susceptible individual is:

$$P_i^{g,s}(t) = 1 - (1 - \rho_A)^{x_i^{g,s}} (1 - \rho_I)^{y_i^{g,s}}. \quad (18)$$

The number of contacts between susceptible individuals of age-group g and asymptomatic individuals in patch i is:

$$x_i^{g,s} = \sum_{f=1}^{N_g} C_{gf} \frac{A_i^{f,s*}}{n_i^{f,s*}} \quad (19)$$

where $A_i^{f,s*}$ is the effective number of asymptomatic individuals of age-group f in patch i at time-slot s , while $n_i^{f,s*}$ is the effective population of age-group f in patch i at time-slot s .

In the same manner, the number of contacts between a susceptible individual of age-group g and symptomatic individuals in patch i is:

$$y_i^{g,s} = \sum_{f=1}^{N_g} C_{gf} \frac{I_i^{f,s*}}{n_i^{f,s*}} \quad (20)$$

Where $I_i^{f,s*}$ is the effective number of symptomatic individuals of age-group f in patch i at time-slot s . The effective populations are given by:

$$\begin{aligned} A_i^{f,s*} &= \sum_{j=1}^N A_{j \rightarrow i}^{f,s} & \text{and} & & A_{j \rightarrow i}^{f,s} &= (r_i^{f,s} \delta_{ij} + p_j^{f,s} M_{ji}^{f,s}) A_j^f(t) \\ I_i^{f,s*} &= \sum_{j=1}^N I_{j \rightarrow i}^{f,s} & \text{and} & & I_{j \rightarrow i}^{f,s} &= (r_i^{f,s} \delta_{ij} + p_j^{f,s} M_{ji}^{f,s}) I_j^f(t) \\ n_i^{f,s*} &= \sum_{j=1}^N n_{j \rightarrow i}^{f,s} & \text{and} & & n_{j \rightarrow i}^{f,s} &= (r_i^{f,s} \delta_{ij} + p_j^{f,s} M_{ji}^{f,s}) n_j^f(t) \end{aligned} \quad (21)$$

where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ otherwise. The effective population in patch j of age-group f is given by:

$$n_j^f(t) = n_{j_0}^f - D_j^f(t) \quad (22)$$

Replacing expressions 19 and 20 in Equation 18, we get:

$$P_i^{g,s}(t) = 1 - \prod_{f=1}^{N_g} \prod_{j=1}^N (1 - \rho_A) C_{gf} \frac{A_{j \rightarrow i}^{f,s}}{n_{j \rightarrow i}^{f,s}} (1 - \rho_I) C_{gf} \frac{I_{j \rightarrow i}^{f,s}}{n_{j \rightarrow i}^{f,s}} \quad (23)$$

Similarly, assuming that all individuals within a household mix homogeneously, the probability of infection within the household is defined as:

$$P_{i(H)}^{g,s}(t) = 1 - (1 - \rho_A)^{(N_{i(H)}-1)\frac{A^{s*}_{i(H)}}{n^{s*}_{i(H)}}} (1 - \rho_I)^{(N_{i(H)}-1)\frac{I^{s*}_{i(H)}}{n^{s*}_{i(H)}}} \quad (24)$$

where $N_{i(H)}$ is the average number of individuals per household in patch i and, $A^{s*}_{i(H)}$ and $I^{s*}_{i(H)}$ are the effective number of asymptomatic, symptomatic, and total number of individuals that stayed at their household in patch i at time-slot s , respectively, given by:

$$\begin{aligned} A^{s*}_{i(H)} &= \sum_{g=1}^{N_g} q_i^{g,s} A_i^g(t) \\ I^{s*}_{i(H)} &= \sum_{g=1}^{N_g} q_i^{g,s} I_i^g(t) \\ n^{s*}_{i(H)} &= \sum_{g=1}^{N_g} q_i^{g,s} n_i^g(t) \end{aligned} \quad (25)$$

Thus, Equation 24 can be written as:

$$P_{i(H)}^{g,s}(t) = 1 - \prod_{g=1}^{N_g} (1 - \rho_A)^{(N_{i(H)}-1)\frac{A_i^{g,s}}{n_i^{g,s}}} (1 - \rho_I)^{(N_{i(H)}-1)\frac{I_i^{g,s}}{n_i^{g,s}}} \quad (26)$$

3.3 Dynamics and stability analysis

For each sub-population for $i \in \{1, \dots, N\}$ and $g \in \{1, \dots, N_g\}$ system 1-8 describes the temporal evolution of the individuals of age-group g in patch i . For the sake of simplicity, the existence, uniqueness, non-negativity and boundedness of the solutions of System 1-8 will be demonstrated for a general sub-population of n individuals. The same applies for the $N \cdot N_g$ subpopulations in the system. Thus, system 1-8 for a general sub-population could be written as:

$$\dot{\vec{X}}(t) = B(t)\vec{X}(t) \quad (27)$$

where $\vec{X}(t) = (S(t), E(t), A(t), I(t), H(t), U(t), R(t), D(t))^T$, $n = \sum_{j=1}^8 X_j(t)$

and $B(t) =$

$$\begin{pmatrix} -\Pi(t) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Pi(t) & -r_{EA} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & r_{EA} & -\alpha r_{AI} - (1-\alpha)r_{AR} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha^9 r_{AI} & -\psi_H r_{IH}^9 - \psi_D r_{ID} - \psi_R r_{IR} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_H r_{IH} & -\eta_U r_{HU} - \eta_D r_{HD} - \eta_R r_{HR} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta_U r_{HU} & -\mu r_{UR} - (1-\mu)r_{UR} & 0 & 0 & 0 \\ 0 & 0 & (1-\alpha)r_{AR} & \psi_R r_{IR} & \eta_R r_{HR} & (1-\mu)r_{UR} & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_D r_{ID} & \eta_D r_{HD} & \mu_U r_{UD} & 0 & 0 & 0 \end{pmatrix}$$

3.3.1 Existence and uniqueness

Theorem 3.1. Cauchy-Lipschitz. Consider the differential equation $x' = f(x)$, $x \in \mathbb{R}^n$, and suppose $f \in C^1$, then exists a unique solution of $x' = f(x)$ with $x(t_0) = x_0$, $x_0 \in \mathbb{R}^n$ and $t_0 \in \mathbb{R}$.

Proposition 3.1. Consider System 27 with non-negative initial conditions, then solution to System 27 exists and is unique for all $t \geq 0$.

Proof. Let $\vec{X}(t) = (S(t), E(t), A(t), I(t), H(t), U(t), R(t), D(t))^T \in \mathbb{R}^8$, then

$$\dot{X}_j(t) = f_j(\vec{X}(t)) \quad (28)$$

where f_j consist of sums of linear terms of $S(t), E(t), A(t), I(t), H(t), U(t), R(t)$ and $D(t)$ for $j \in \{1, \dots, 8\}$. Thus, f_j are continuous functions in $\mathbb{R}^8$ and $\frac{dX_j(t)}{dt}$ exist. Then, according to Theorem 1, exists a unique solution of Equations 28 for any initial condition. $\square$

3.3.2 Non-negativity

Proposition 3.2. Given non-negative initial conditions of System 27, the solutions of system 27, $S(t), E(t), A(t), I(t), H(t), U(t), R(t)$ and $D(t)$, are nonnegative for all $t \geq 0$.

Proof. Let $t_s \geq 0$ be so that $S(t_s) = 0$ and $X_j(t_s) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 1$. Thus, from System 27 we get:

$$S'(t_s) = 0$$

Let $t_e \geq 0$ be so that $E(t_e) = 0$ and $X_j(t_e) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 2$. Thus, from System 27 we get:

$$E'(t_e) = \Pi(t_e) S(t_e) \geq 0$$

given that the probability $\Pi(t) \in [0, 1]$.

Let $t_A \geq 0$ be so that $A(t_A) = 0$ and $X_j(t_A) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 3$. Thus, from System 27 we get:

$$A'(t_A) = r_{EA} E(t_A) \geq 0$$

Let $t_I \geq 0$ be so that $I(t_I) = 0$ and $X_j(t_I) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 4$. Thus, from System 27 we get:

$$I'(t_I) = \alpha r_{AI} A(t_I) \geq 0$$

Let $t_H \geq 0$ be so that $H(t_H) = 0$ and $X_j(t_H) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 5$. Thus, from System 27 we get:

$$H'(t_H) = \psi_H r_{IH} I(t_H) \geq 0$$

Let $t_U \geq 0$ be so that $U(t_U) = 0$ and $X_j(t_U) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 6$. Thus, from System 27 we get:

$$U'(t_U) = \eta_U r_{HU} H(t_U) \geq 0$$

Let $t_R \geq 0$ be so that $R(t_R) = 0$ and $X_j(t_R) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 7$. Thus, from System 27 we get:

$$R'(t_R) = (1-\alpha)r_{AR} A(t_R) + \psi_R r_{IR} I(t_R) + \eta_R r_{HR} H(t_R) + (1-\mu)r_{UR} U(t_R) \geq 0$$

Let $t_D \geq 0$ be so that $D(t_D) = 0$ and $X_j(t_D) \geq 0$ for $j \in \{1, \dots, 8\}$ and $j \neq 8$. Thus, from System 27 we get:

$$D'(t_D) = \psi_D r_{ID} I(t_D) + \eta_D r_{HD} H(t_D) + \mu r_{UD} U(t_D) \geq 0$$

Hence, none of the states will have a negative derivative at any time-instant when they vanish and the rest of the states are non negative. Then, all states must have non-negative values for all time all $t \geq 0$.

QED.

3.3.3 Boundedness

Summing all the equations in System 1-8 gives the evolution of the total population $M(t)$:

$$M'(t) = 0 \rightarrow M(t) \text{ constant} \quad (29)$$

Thus, System 1-8 is well posed since:

- Solutions exist and are unique
- Remain non-negative for non-negative initial conditions
- total population size is constant

4 Numerical simulation of the Covid-19 spreading

The ODE based epidemic model presented in Section 3.1 is used to simulate the epidemic spreading of COVID-19. Note that the deterministic epidemic models are a good approximation of the epidemic dynamics only for large populations [30]. In order to study the effect of few infectious individuals in the epidemic dynamics one should include a probability distribution (such as the probability

of infection described in 3.2, for example). The system of differential equations is solved numerically using Isoda from the FORTRAN library odepack.

To check the validity of our model, we first simulate the epidemic spreading in a single population without mobility.

4.1 Single population

We consider a population of $N=101$ people, with 1 infected individual and 100 susceptible individuals, $(S_0, E_0, A_0, I_0, H_0, U_0, R_0, D_0) = (100, 0, 0, 1, 0, 0, 0, 0)$. There is not mobility between patches. For shake of simplicity, we consider a single age-range and a single time-slot. A summary of the parameters considered in the simulation are given in Table 1.

We simulate the evolution of the disease over 100 days and we compare the results obtained considering homogeneous mixing and contact based mixing.

Parameter	Value	Parameter	Value
N	1	r_{UD}	0.14 days ⁻¹
N_g, N_h	1	r_{UR}	0.1 days ⁻¹
C_{11}	10		
$\{p, q, r\}$	$\{0, 0, 1\}$	α	0.8
r_{EA}	0.29 days ⁻¹	ψ_H	0.38
r_{AI}	0.4 days ⁻¹	ψ_D	0.02
r_{AR}	0.35 days ⁻¹	ψ_R	0.6
r_{IH}	0.17 days ⁻¹	η_U	0.16
r_{ID}	0.06 days ⁻¹	η_D	0.02
r_{IR}	0.33 days ⁻¹	η_R	0.76
r_{HU}	0.5 days ⁻¹	μ	0.18
r_{HD}	0.11 days ⁻¹	ρ_I	0.06
r_{HR}	0.13 days ⁻¹	ρ_A	0.03

Table 1. List of parameters used to create the baseline model.

Under the assumption of homogeneous mixing, the infection rate of infectious individuals is $\beta_i = 0.62$ calculated by Eq. 9 with $\rho_i = 0.06$. The infectivity of asymptomatic individuals is considered to be half of one of infected individuals.

Under the contact based mixing assumption, we assume that the average number of contacts between individuals per day is $C_{11} = 10$. The probability of infection of infected and asymptomatic individuals is given in Table 1.

The evolution of the population is shown in Figure 4b for homogeneous mixing and in Figure 4b for contact-based mixing.

Both models reproduce similar results. In cumulative numbers, the error (difference between the two models predictions) in the predicted number of individuals in each compartment ranges from 0.03 to 0.7 individuals. Figure 5 shows the cumulative and predicted daily new cases for both mixing models.

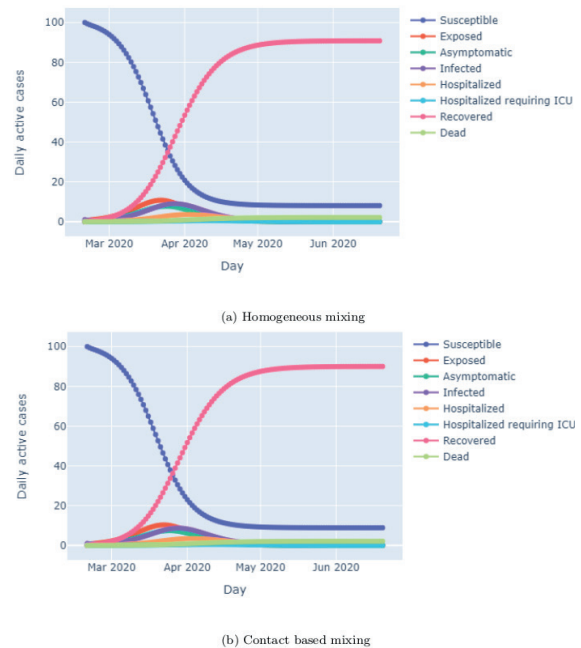


Figure 4. Simulation of the epidemic model - Cases per day

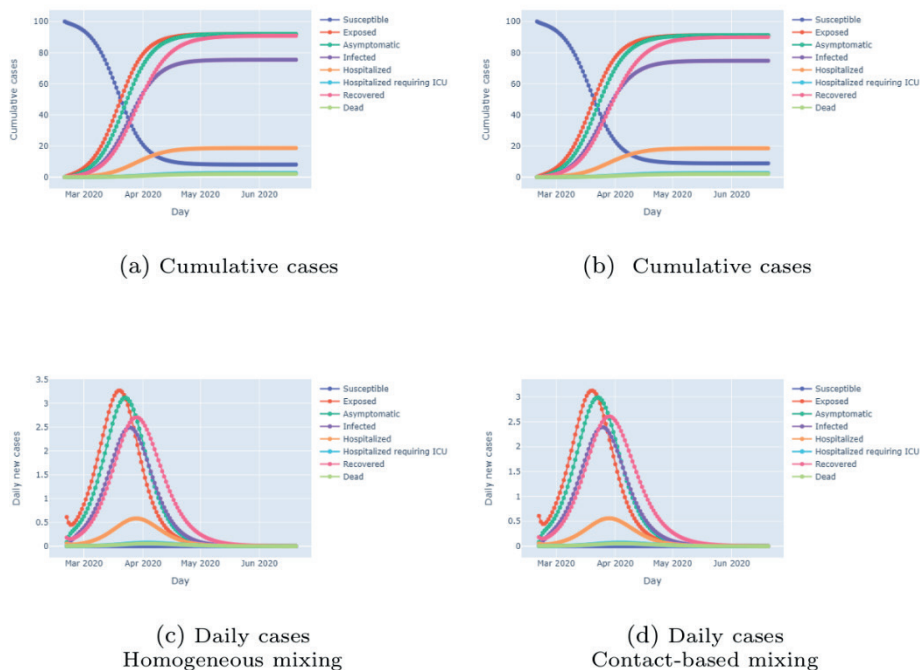


Figure 5. Cumulative (up) and daily new cases (down) for the homogeneous (left) and contact-based (right) mixing models.

4.2 Spain with mobility

Mobility data is taken from the Instituto Nacional de Estadística (INE) [28]. The Spanish region is divided into 3214 mobility areas. Mobility areas are composed by municipalities of the same province but for the ones in Table 2 which belong to a different province than the one of the

mobility region. Simulation results are given by mobility region. Thus, note that results aggregated by province will show a small error due to the fact that the cases of those municipalities will be allocated to a different province than the one they geographically belong to.

Municipality	Province	Mobility region	Mobility Province
Valle de Villaverde	Cantabria	01BI	Bizkaia
Petilla de Aragón	Navarra	01ZA	Zaragoza
Villodrigo	Palencia	12BU	Burgos

Table 2: Municipalities considered inside a mobility province which does not correspond to their geographical province.

Data is offered from March 16 every other day and from March 31, daily data, until June 16. Mobility data is described by mobility and displacement data. The **displacement data** shows the **daily number of individuals** moving in their patch or moving to other patches and the number of destinations. The **mobility data** shows the **number of trips observed** (number of individuals travelling) between patches by day if greater or equal to 100. The destination of trips (fluxes) smaller than 100 is not given. This leads to a number of people leaving their patch to unknown destinations. In the absence of mobility data for the date to be simulated, the average mobility values of the corresponding weekday are considered.

For example, the 1st of April, 898 persons move from patch '001A' to 56 destinations. From patch '001A' 3 known destinations are given with a total flux of 425 people. Thus, there are $898 - 425 = 473$ people leaving patch '001A' to $56 - 3 = 53$ remaining unknown destinations, with fluxes below 100.

We considered two options to overcome the issue of people leaving their patch to unknown destinations.

1. **Option 1:** consider only the trips with known destinations, i.e. consider only mobility data.

Let n_i be the base population of patch i . Let n_{ij} be the number of individuals travelling from patch i to patch j by day given by the mobility data. Thus, displacement and mobility rates for each day are calculated as follows:

$$p_i = \frac{\sum_{j=1}^N n_{ij}}{n_{i0}} \quad (30)$$

$$r_i = 1 - p_i \quad (31)$$

$$M_{ij} = \frac{n_{ij}}{\sum_{j=1}^N n_{ij}} \quad (32)$$

2. **Option 2:** allocate randomly the remaining people to the remaining number of closest destinations.

In this example, the first option is considered to simulate the spread of coronavirus in Spain. We simulate the evolution of the disease from February to July under contact based mixing assumption and considering the parameters shown in Table 1. The evolution of the population is shown in Figure 6

Note that the peak is simulated around end of May, which does not correspond to the first outbreak of Spain (peak was observed around March-April). Hence, the parameter estimations considered do not match the observations. For this reason, a parameter inference needs to be performed.

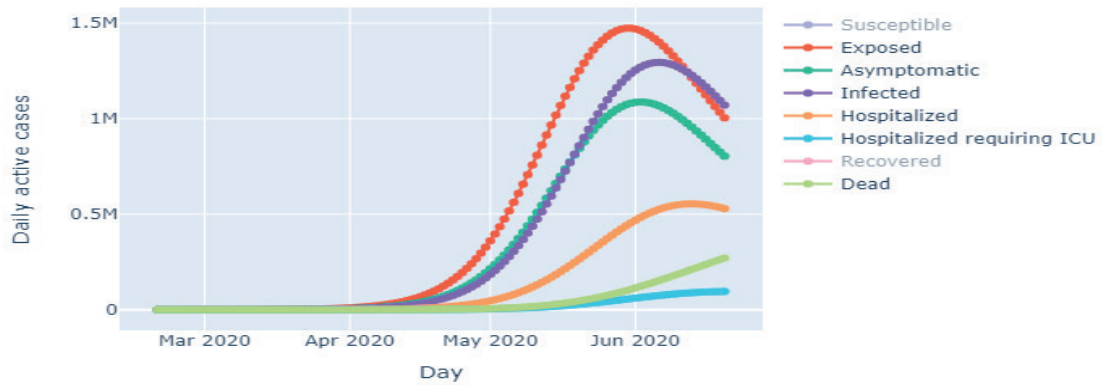


Figure 6. Epidemiological spread in Spain - Cases per day

5 Calibration/Parameter inference

The model is calibrated by adjusting the model parameters to obtain the best fit between the model calculations and the observed data. We apply Approximate Bayesian Computation method based on sequential Monte Carlo (ABC-SMC) to estimate the parameters of the dynamical

model (their posterior distribution). The most common summary statistics for posterior distributions are the posterior mean, posterior mode and posterior median. Often, a safe assumption is to report the posterior mean and posterior standard deviation [36].

5.1 Approximate Bayesian Computation based on sequential Monte Carlo

We use the pyABC python package for Approximate Bayesian Computation (ABC), using a Sequential Monte Carlo scheme. The ABC algorithm (SMC). The fundamental components of the algorithm are the following:

- Observed data y_0
- Simulation based model $M(y^o, \theta)$
- Prior distribution of the parameters $\varphi(\cdot)$
- Summary statistics
- Distance or likelihood function: function to compute approximate likelihood function based on the extracted summary statistics
- Kernel: defines the exploration of the parameter space
- A particle is a sample of parameters θ from the prior distribution. A population is a sample of N accepted particles.
- Tolerance : particle accepted if the distance between observed and simulated data is smaller than the given tolerance.

A detailed description of the algorithm is given in [37].

5.2 Parameter inference of the ODE based model

5.2.1 Optimal ABC-SMC algorithm settings

In order to perform the parameters inference we first need to define the settings of ABC-SMC algorithm.

- Observed data: The model is calibrated to fit the number of deaths observed within the simulated period.
- Parameter prior distribution: two prior-distributions have been considered: uniform distribution in $[0,1]$ and adjusted uniform distribution which considers the observed value of the parameters.

The distance function which best measures the error between sampled and observed data. Here we defined the distance function as:

$$d(y_0, y_{sim}) = \frac{\sum |y_0 - y_{sim}|}{n} \quad (33)$$

where y_{sim} is the simulated data and n the number of points observed (length of vectors).

- Kernel: Transition via a multivariate Gaussian KDE estimate. MultivariateNormalTransition with scaling parameter equal to 1.
- Epsilon sampling strategy: Median Epsilon calculates epsilon as median of the distances from the last population.
- Number of particles N per population to estimate
- the posterior distributions of the parameters. We set $N = 1000$.
- Maximum number of populations T_{max} . We defined $T_{max} = 15$
- Tolerance, ϵ .

The use-case of Spain without mobility is used to find the optimal algorithm settings (prior distribution and tolerance) which lead to the best algorithm performance in terms of best parameter inference and smaller computation time. The epidemic spread in Spain is simulated from the 20st of February 2020 to 21st of May 2020 (Spanish first wave). The number of deaths reported in this period are used to calibrate the model. Table 3 resumes the different calibration scenarios (ABC-SMC algorithm settings) considered and the error obtained. The error is given as the distance between observed deaths and simulated deaths. The mean sample (simulated deaths from the mean parameter values of last accepted population) and the best sample (best approximation to the observed deaths of the last accepted population) are taken into consideration. Empty lines corresponds to presently running experiments.

DB-run	Priors	ϵ	Execution-time	Error(mean)	Error(Best fit)
1-5	uniform	1e-04	4.5 h	51.12%	6.53%
2-1	adjusted	1e-04	36 h	56.35%	1.79%
1-7	uniform	1e-05			
3-1	adjusted	1e-05			
2-2	uniform	1e-06			
3-2	adjusted	1e-06			

Table 3. Calibration algorithm settings and results

Figure 7 shows the mean sample and best fitting sample obtained from the calibrations compared to the observed deaths. Thus far, calibration 2-1 best fitting sample shows the best results, with a mean error of 1,79%, i.e. a mean difference of 500 deaths.

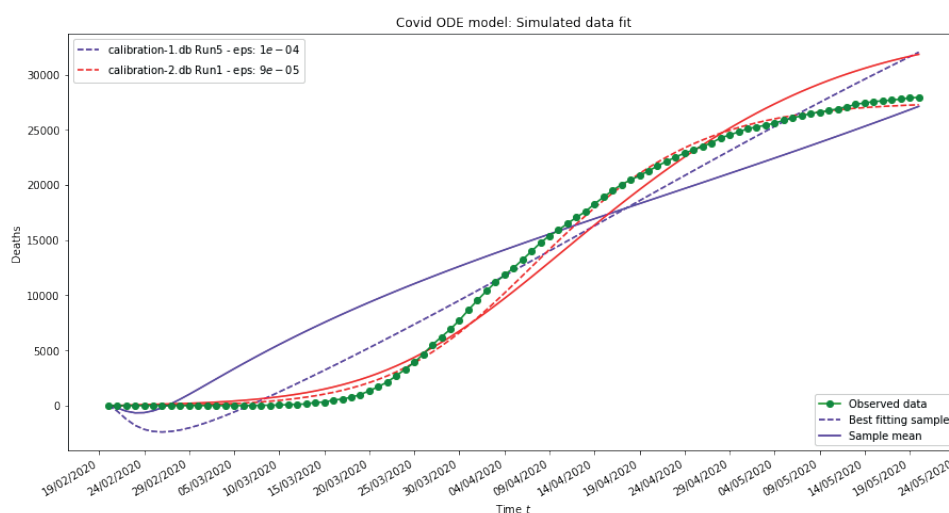


Figure 7. Calibration results - Number of deaths

Figure 8 shows the error with respect to the observed values.

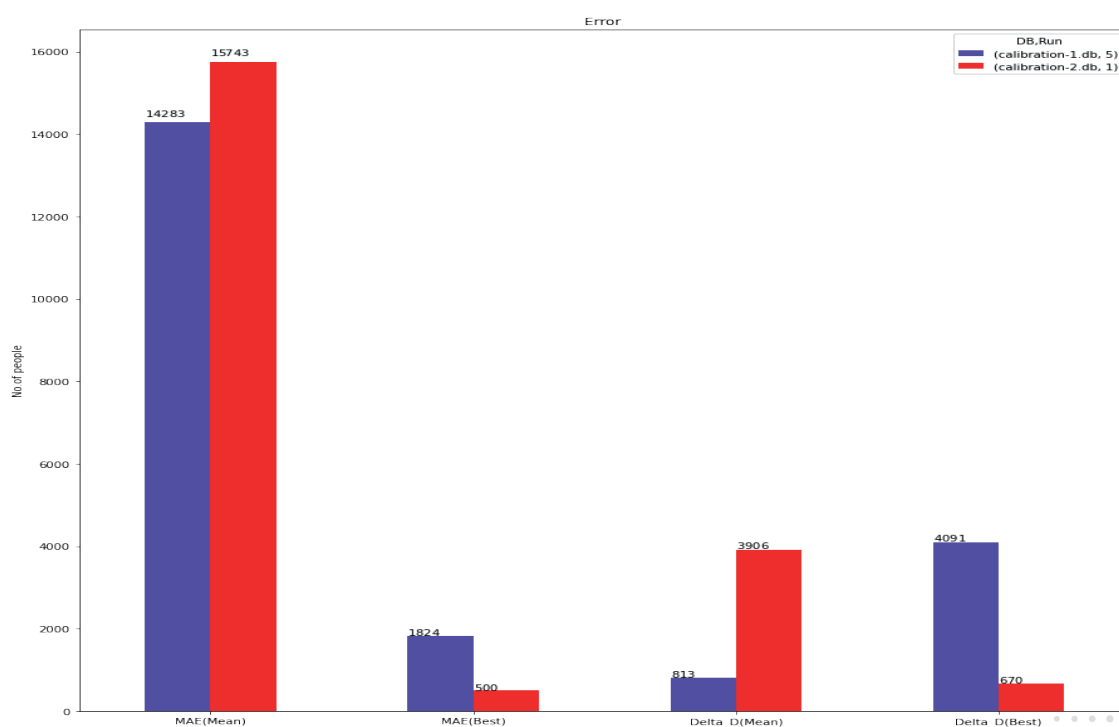


Figure 8. Calibration results - Error

Figure 9 shows the posterior mean and posterior standard deviation of the model parameters compared to the preset values.

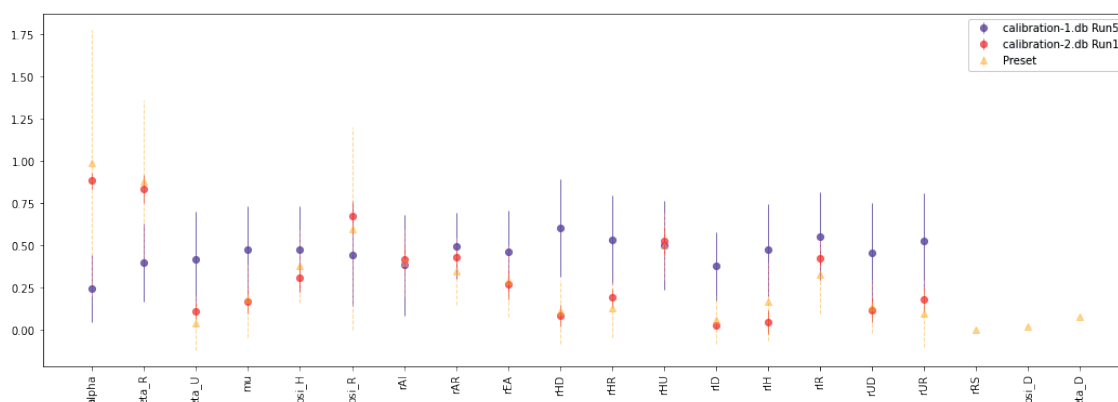


Figure 9. Posterior mean and posterior standard deviation of model parameters

The following sections consider Calibration 2-1 results.

Parameter	Value
r_{EA}	0.4 days ⁻¹
r_{AI}	0.46 days ⁻¹
r_{AR}	0.5 days ⁻¹
r_{IH}	0.005 days ⁻¹
r_{ID}	0.0004 days ⁻¹
r_{IR}	0.39 days ⁻¹
r_{HU}	0.53 days ⁻¹
r_{HD}	0.14 days ⁻¹
r_{HR}	0.25 days ⁻¹

Parameter	Value
r_{UD}	0.16 days ⁻¹
r_{UR}	0.16 days ⁻¹
α	0.85
ψ_H	0.28
ψ_D	0.02
ψ_R	0.7
η_U	0.15
η_D	0.1
η_R	0.75
μ	0.31

Table 4. Best fitting sample - Inferred parameter values

5.2.2 Best fitting sample: Posterior parameter values and prediction

Posterior parameter values of the best fitting sample of Calibration 2-1 are given in Table 5.2.2. Initial conditions are $I=73$ infected individuals and $A=230511$ asymptomatic individuals. The total population in Spain is 47026135 people. All other parameter values are considered as in Table 1.

The number of deaths simulated and observed are compared in Figure 10.

Considering the inferred parameter values of the best sample, the daily new cases simulated are shown in Fig 11.

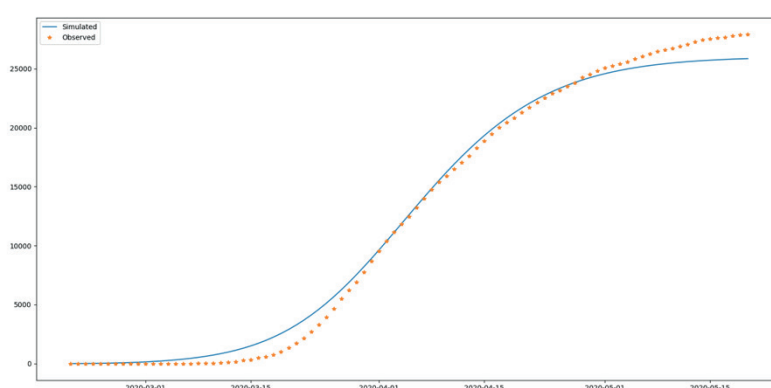
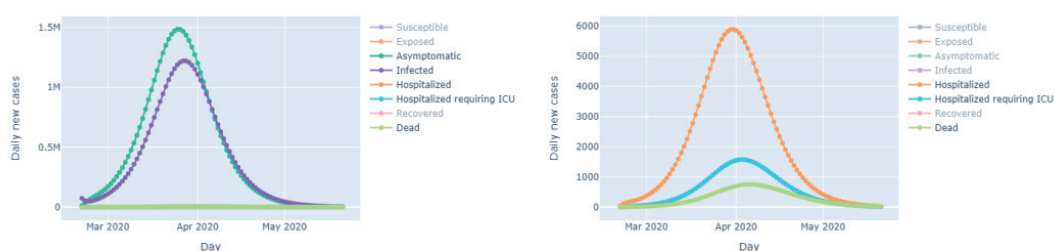


Figure 10. Best fitting sample - Number of deaths



(a) Asymptomatic and Infected

(b) Hospitalized , ICU and deaths

Figure 11. Daily new cases: Asymptomatic and Infected (left) and Hospitalized, ICU and deaths (right)

The comparison between simulated cases and observed cases is shown in Fig. 12. While the best sample shows a good fit in number of deaths, the number of Hospitalized and ICU cases is not well approximated. In fact, although the trend is correct, the model simulates many more cases than observed.

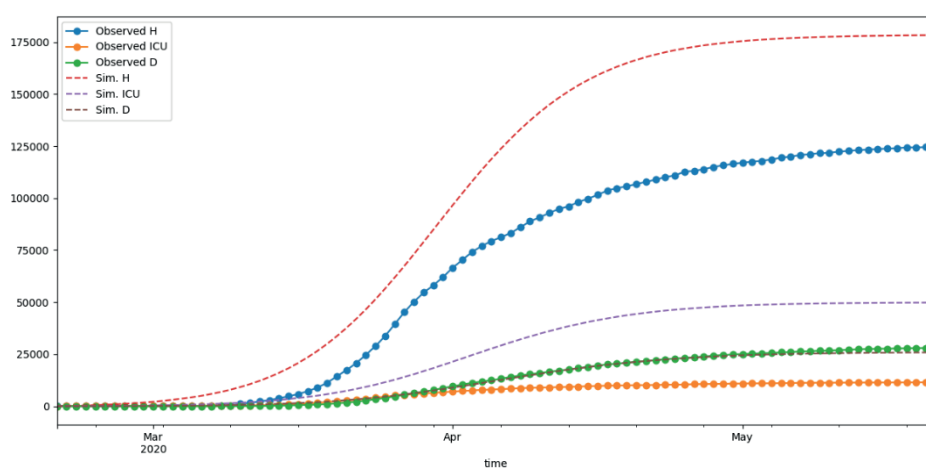


Figure 12. Best fitting sample compared to observed cases- Cumulative Hospitalized, ICU and deaths cases

As explained before and in order to obtain better estimations, many other calibrations are presently being executed.

References

1. WHO, "Novel coronavirus—china," 2020.
2. W. H. Organization *et al.*, "Who director-general's remarks at the media briefing on 2019-ncov on 11 february 2020," *World Health Organization, Geneva*. Available via <https://www.who.int/dg/speeches/detail/whodirector-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11february-2020>. Accessed, vol. 10, 2020.
3. P. H. England, "Covid-19: epidemiology, virology and clinical features," 2020.
4. P. Linde, "Sanidad confirma en La Gomera el primer caso de coronavirus en España," *El País*, vol. 1, 2020.
5. E. COVID-19, RENAVE, CNE, and C. (ISCIII), "Informe sobre la situación de covid-19 en España," Informe general 29, Red Nacional de Vigilancia Epidemiológica (RENAVE), May 2020. Date of the report: 07-05-2020.
6. Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, *et al.*, "Epidemiology and transmission of covid-19 in shenzhen china: Analysis of 391 cases and 1,286 of their close contacts," *MedRxiv*, 2020.
7. M. de Sanidad, "Consumo y bienestar social. información científico-técnica, enfermedad por coronavirus, covid-19," 2020.
8. J. F.-W. Chan, S. Yuan, A. J. Zhang, V. K.-M. Poon, C. C.-S. Chan, A. C.Y. Lee, Z. Fan, C. Li, R. Liang, J. Cao, *et al.*, "Surgical mask partition reduces the risk of non-contact transmission in a golden syrian hamster model for coronavirus disease 2019 (covid-19)," *Clinical Infectious Diseases*, 2020.
9. J. Lu, J. Gu, K. Li, C. Xu, W. Su, Z. Lai, D. Zhou, C. Yu, B. Xu, and Z. Yang, "Covid-19 outbreak associated with air conditioning in restaurant, guangzhou, china, 2020," *Emerging infectious diseases*, vol. 26, no. 7, p. 1628, 2020.
10. S. Asadi, A. S. Wexler, C. D. Cappa, S. Barreda, N. M. Bouvier, and W. D. Ristenpart, "Aerosol emission and superemission during human speech increase with voice loudness," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.
11. S. L. Miller, W. W. Nazaroff, J. L. Jimenez, A. Boerstra, G. Buonanno, S. J. Dancer, J. Kurnitski, L. C. Marr, L. Morawska, and C. Noakes, "Transmission of sars-cov-2 by inhalation of respiratory aerosol in the skagit valley chorale superspreading event," *Indoor air*, 2020.
12. J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, *et al.*, "Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *lancet glob, Health*, vol. 8, no. 4, pp. 488–496, 2020.
13. K. E. Nelson and C. M. Williams, *Infectious disease epidemiology: theory and practice*. Jones & Bartlett Publishers, 2014.
14. X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, *et al.*, "Temporal dynamics in viral shedding and transmissibility of covid-19," *Nature medicine*, pp. 1–4, 2020.
15. A. Grifoni, D. Weiskopf, S. I. Ramirez, J. Mateus, J. M. Dan, C. R. Moderbacher, S. A. Rawlings, A. Sutherland, L. Premkumar, R. S. Jodi, *et al.*, "Targets of t cell responses to sars-cov-2 coronavirus in humans with covid19 disease and unexposed individuals," *Cell*, 2020.
16. R. Woelfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Mueller, D. Niemeyer, P. Vollmar, C. Rothe, M. Hoelscher, *et al.*, "Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster," *MedRxiv*, 2020.
17. J. Zhao, Q. Yuan, H. Wang, W. Liu, X. Liao, Y. Su, X. Wang, J. Yuan, T. Li, J. Li, *et al.*, "Antibody responses to sars-cov-2 in patients of novel coronavirus disease 2019," *Clinical Infectious Diseases*, 2020.
18. N. M. Okba, M. A. Muller, W. Li, C. Wang, C. H. Geurts-vanKessel, V. M. Corman, M. M. Lamers, R. S. Sikkema, E. de Bruin, F. D. Chandler, *et al.*, "Sars-cov-2 specific antibody responses in covid-19 patients," *medRxiv*, 2020.
19. W. Liu, L. Liu, G. Kou, Y. Zheng, *et al.*, "Evaluation of nucleocapsid and spike protein-based elisas for detecting antibodies against sars-cov-2. medrxiv [internet]. 2020."
20. Q.-x. Long, H.-j. Deng, J. Chen, J. Hu, B.-z. Liu, P. Liao, Y. Lin, L.-h. Yu, Z. Mo, Y.-y. Xu, *et al.*, "Antibody responses to sars-cov-2 in covid-19 patients: the perspective application of serological tests in clinical practice," *medRxiv*, 2020.
21. W. Y. Wan, S. H. Lim, and E. H. Seng, "Cross-reaction of sera from covid19 patients with sars-cov assays," *medRxiv*, 2020.

22. L.-P. Wu, N.-C. Wang, Y.-H. Chang, X.-Y. Tian, D.-Y. Na, L.-Y. Zhang, L. Zheng, T. Lan, L.-F. Wang, and G.-D. Liang, "Duration of antibody responses after severe acute respiratory syndrome," *Emerging infectious diseases*, vol. 13, no. 10, p. 1562, 2007.
23. P. Kellam and W. Barclay, "The dynamics of humoral immune responses following sars-cov-2 infection and the potential for reinfection," *Journal of General Virology*, p. jgv001439, 2020.
24. A. T. Xiao, C. Gao, and S. Zhang, "Profile of specific antibodies to sarscov-2: the first report," *The Journal of infection*, 2020.
25. H. Zeng, C. Xu, J. Fan, Y. Tang, Q. Deng, W. Zhang, and X. Long, "Antibodies in infants born to mothers with covid-19 pneumonia," *Jama*, vol. 323, no. 18, pp. 1848–1849, 2020.
26. K. Callow, H. Parry, M. Sergeant, and D. Tyrrell, "The time course of the immune response to experimental coronavirus infection of man," *Epidemiology & Infection*, vol. 105, no. 2, pp. 435–446, 1990.
27. S. M. Iacus, C. Santamaria, F. Sermi, S. Spyrtos, D. Tarchi, and M. Vespe, "How human mobility explains the initial spread of covid-19," *Publications Office of the European Union*, 2020.
28. https://www.ine.es/covid/covid_ovilidad.htm.
29. D. Soriano-Paños, G. Ghoshal, A. Arenas, and J. Gómez-Gardeñes, "Impact of temporal scales and recurrent mobility patterns on the unfolding of epidemics," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2020, p. 024006, feb 2020.
30. M. Choisy, J.-F. Guégan, and P. Rohani, *Mathematical Modeling of Infectious Diseases Dynamics*, ch. 22, pp. 379–404. John Wiley Sons, Ltd, 2006.
31. D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, "Spreading processes in multiplex metapopulations containing different mobility networks," *Phys. Rev. X*, vol. 8, p. 031039, Aug 2018.
32. D. Li, Q. Shao, and D. Han, "Exploring the Epidemic Spreading in a Multilayer Metapopulation Network by considering Individuals' Periodic Travelling," *Complexity*, vol. 2020, pp. 1–9, April 2020.
33. C. Granell and P. Mucha, "Epidemic spreading in localized environments with recurrent mobility patterns," *Physical Review E*, vol. 97, 04 2018.
34. S. Gomez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, "Discrete-time markov chain approach to contact-based disease spreading in complex networks," *EPL*, vol. 89, 07 2009.
35. F. Ball, T. Britton, T. House, V. Isham, D. Mollison, L. Pellis, and G. S. Tomba, "Seven challenges for metapopulation models of epidemics, including households models," *Epidemics*, vol. 10, pp. 63 – 67, 2015. Challenges in Modelling Infectious Disease Dynamics.
36. M. Whalen, *Biologists Journey Into Bayesian Inference*. 2020.
37. T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf, "Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems," *Journal of The Royal Society Interface*, vol. 6, p. 187–202, Jul 2008.



