

Quantifying the non-reported new daily cases of COVID-2019 by region in Spain at a real-time

The present outbreak of COVID-19 disease, caused by the SARS-CoV-2 virus, has put the planet in quarantine. On January 30, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a “public health emergency of international concern”, and then a pandemic on March 11.

Spain has become the fourth country worldwide with more infected cases, officially registering over thousands of cases in a short time. Although many critical and severe measures have been considered from the authorities to lessen the impact of the outbreak and help flatten the curve, they rely on numbers that could be unreliable and therefore misrepresent the implications of such pandemic.

Counts in Spain due to the protocols used for testing, mainly include individuals with severe symptoms. The authorities have just announced a new protocol with rapid tests to be implemented in a few days elpais.com.

Given the nature of our data, we can guess that the estimated number of cases that we are finding are in fact potentially severe cases, and presumably the size of the infected population (asymptomatic) is even higher.

Accordingly, the current analysis aims to update the situation concerning COVID-19 daily, and particularly quantify the potential under-reporting in the official registered cases by region in Spain. Results herein can help to have a more realistic picture of the pandemic at a real time as well as to more accurately estimate essential measures such as the basic reproduction number or the fatality rate that are used for practitioners and politicians to make decisions.

The data for the analysis have been extracted from eldiario.es, where official data are gathered.

Notice that this analysis can be easily reproduced for other countries.

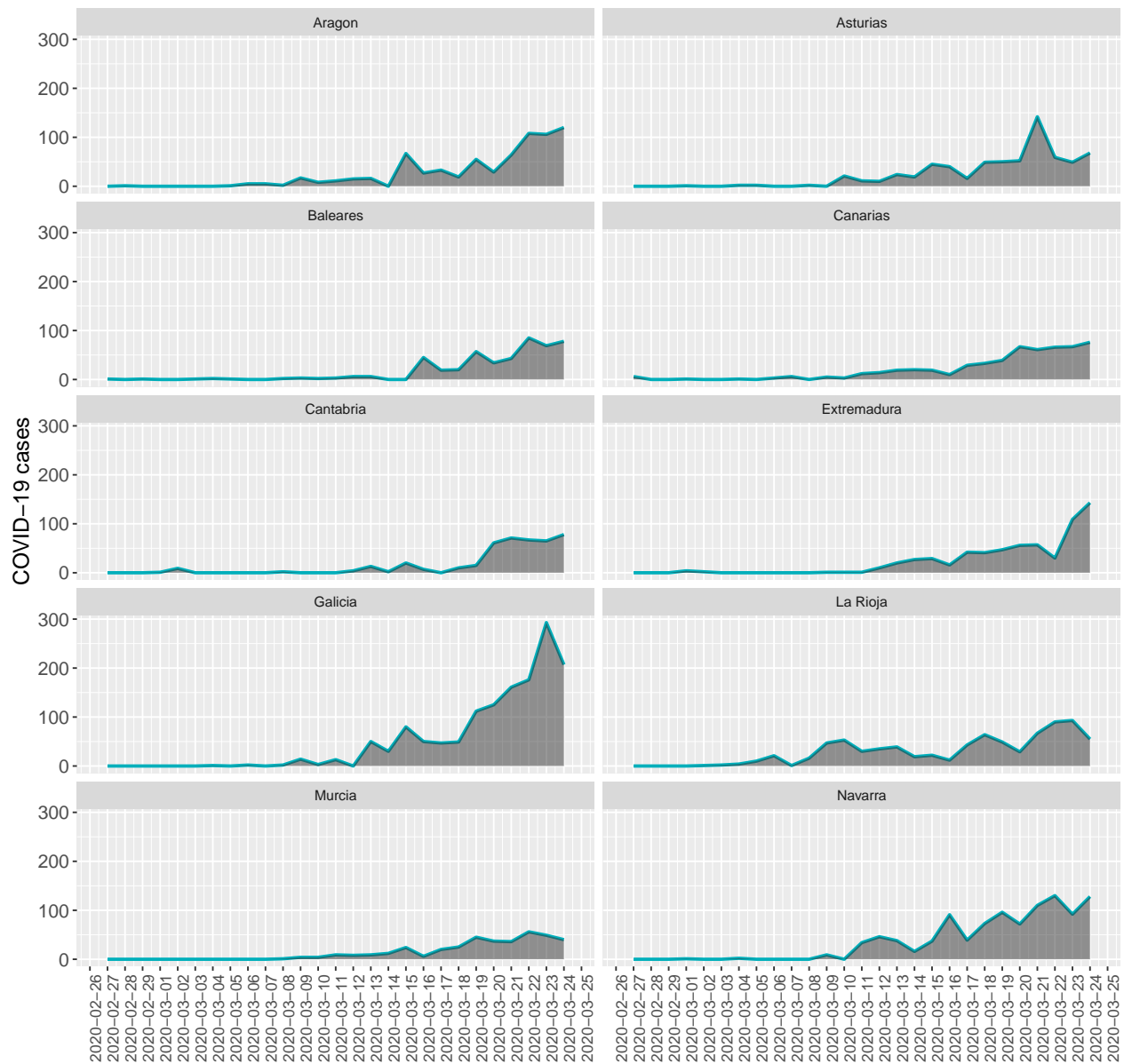


Figure 1 (a): Daily COVID-19 cases from 27-02-2020 to 24-03-2020

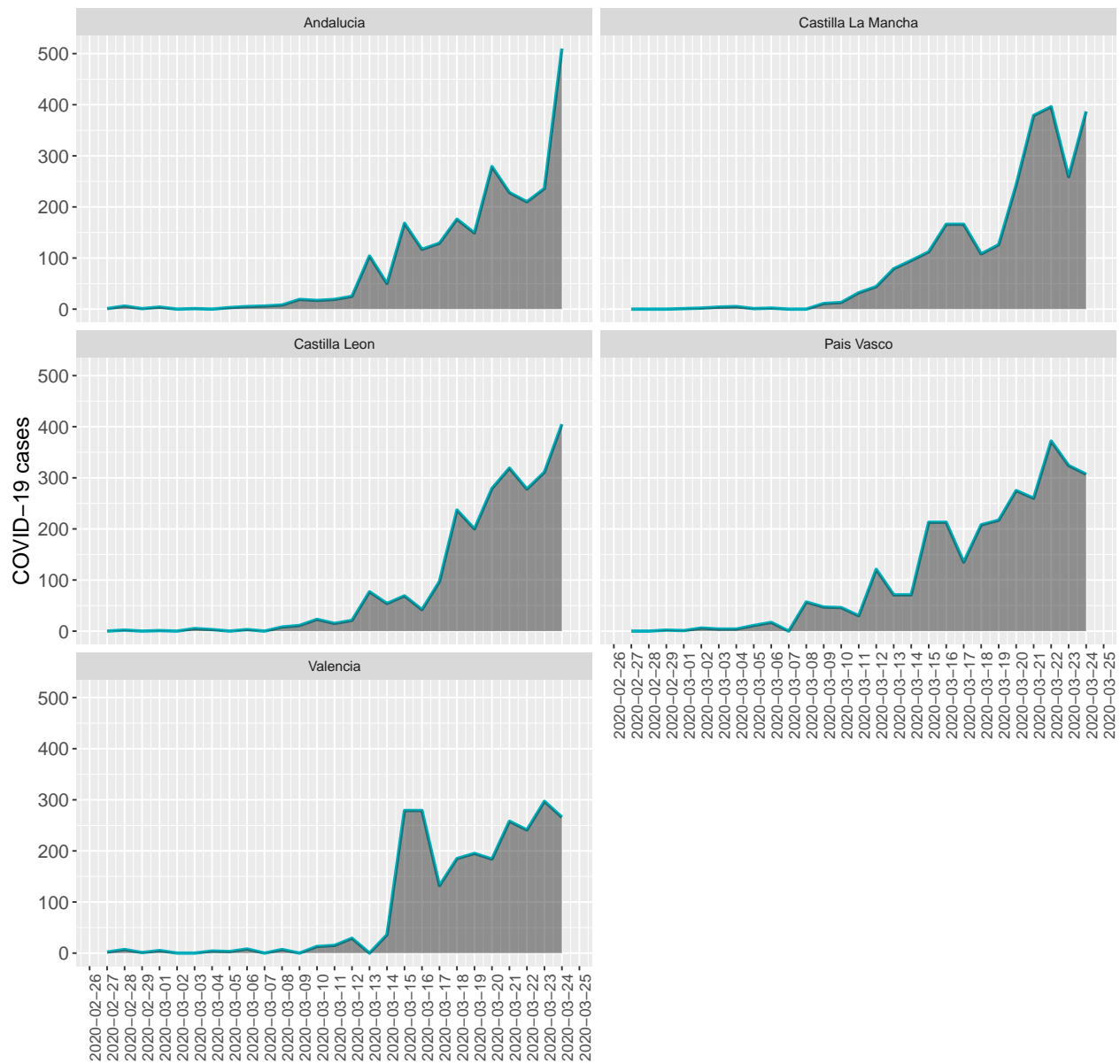


Figure 1 (b): Daily COVID-19 cases from 27-02-2020 to 24-03-2020

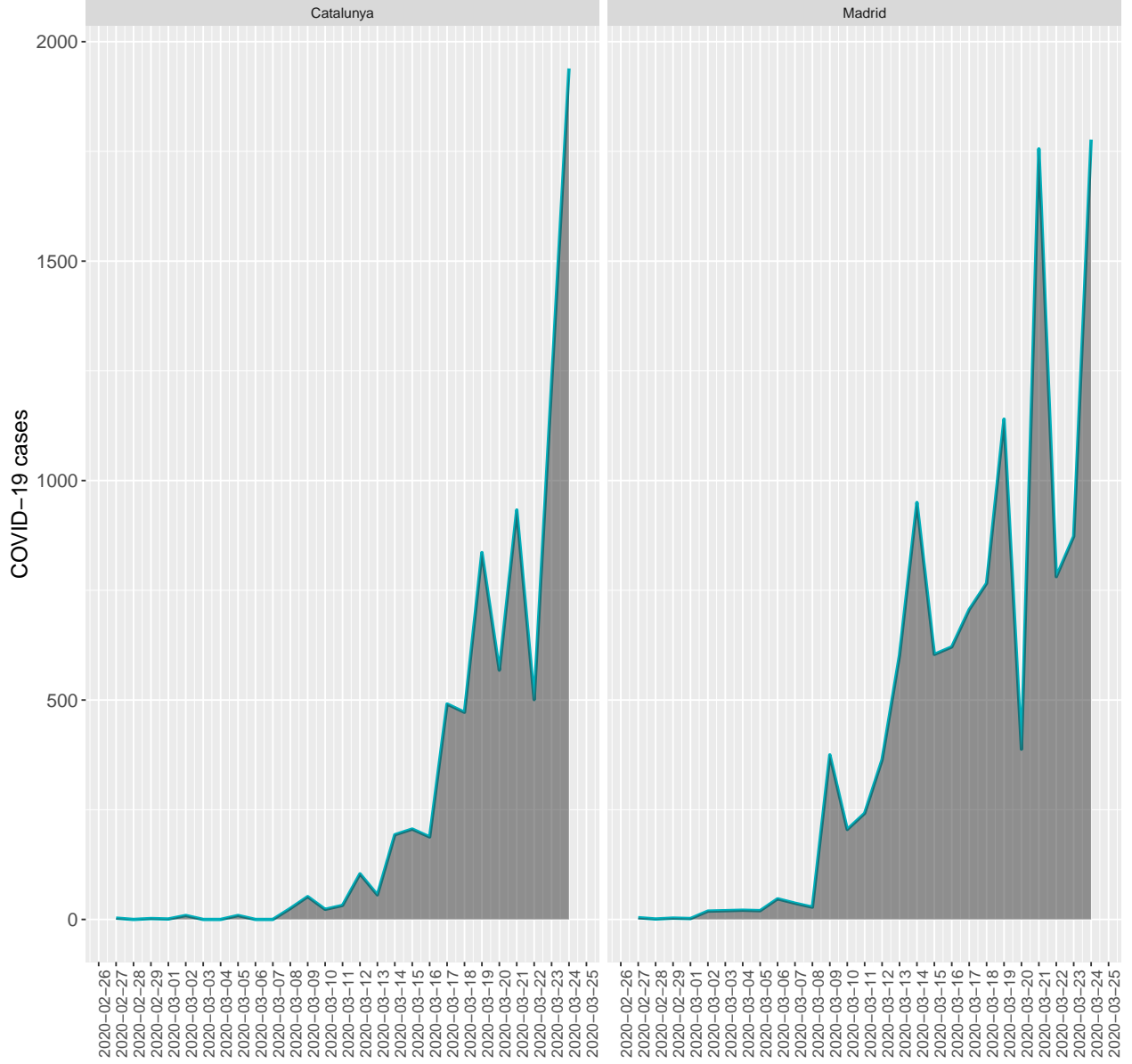


Figure 1 (c): Daily COVID-19 cases from 27-02-2020 to 24-03-2020

If the under-reporting is ignored, the daily counts can be appropriately modeled following: $\exp(\alpha_0 + \alpha_1 t)$, since the number of daily COVID-19 cases overtime properly grows exponentially according to Figure 1.

However, if we consider that the official number of daily cases does not reflect the total number of cases (e.g., a proportion of the cases is not observed, and thus the data are misreported), the model above does not make any sense, and therefore a more appropriate alternative should be considered.

We shall base all the subsequent analysis in a model introduced by Fernández-Fontelo et al. (2016).

In that model, two different processes are considered: X_n which is the true process but unobserved (latent), and Y_n which is observed and potentially under-reported. In this application, the latent process is assumed to be Poisson distributed with time-dependent rate, $\lambda_t = \exp(\beta_0 + \beta_1 t)$. The observed process will always be lower or equal than the latent process (due to the under-reporting) in such a way that Y_n will be equal than X_n (non under-reporting) with probability $1 - \omega$; or Y_n is $q \circ X_n$ with probability ω . Parameters ω and q quantify the overall frequency and intensity of the phenomenon, which roughly speaking describe respectively the number of times the observed counts are not equal to the real ones, and the distance between the real

	minimum	mean	median	maximum	standard deviation	dispersion index
Andalucia	0.00	91.52	19.00	510.00	122.55	164.09
Aragon	0.00	26.26	11.00	120.00	36.33	50.27
Asturias	0.00	24.52	11.00	142.00	32.54	43.20
Baleares	0.00	17.70	2.00	85.00	26.86	40.74
Canarias	0.00	20.63	10.00	76.00	25.17	30.71
Cantabria	0.00	15.74	2.00	78.00	26.25	43.78
Castilla La Mancha	0.00	97.44	32.00	396.00	129.16	171.20
Castilla Leon	0.00	91.11	21.00	405.00	126.47	175.56
Catalunya	0.00	291.26	52.00	1939.00	468.26	752.83
Extremadura	0.00	23.56	4.00	143.00	35.43	53.29
Galicia	0.00	52.41	13.00	293.00	77.68	115.15
La Rioja	0.00	29.70	22.00	93.00	27.80	26.03
Madrid	1.00	457.48	364.00	1777.00	514.89	579.51
Murcia	0.00	14.26	6.00	56.00	17.95	22.60
Navarra	0.00	37.56	16.00	130.00	44.60	52.97
Pais Vasco	0.00	111.56	57.00	372.00	120.55	130.26
Valencia	0.00	90.59	13.00	297.00	115.02	146.04

Table 1: Summary of the daily COVID-19 cases from 27-02-20 to 24-03-2020 by region in Spain

and observed processes.

Using the Viterbi algorithm, the model also enables reconstructing the most likely sequence of real COVID-19 cases throughout the study. This allows us to have an estimated time series of truly daily cases and evaluate the impact of under-reporting over measures such as the basic reproduction number. Figure 2 shows the observed and reconstructed series over time by region.

	α	β_0	β_1	ω	q	AIC
Andalucia	0.6485	0.2216	0.713	0.5269		330.2
s.e. (Andalucia)	0.1204	0.0056	0.0955	0.024		
Aragon	0.156	0.2124	0.9144	0.347		207.2
s.e. (Aragon)	0.2007	0.008	0.0583	0.0208		
Asturias	-0.5688	0.2229	0.5385	0.3382		196.8
s.e. (Asturias)	0.2916	0.0136	0.1274	0.0333		
Baleares	-0.6082	0.1903	0.0944	0		184.1
s.e. (Baleares)	0.2391	0.0102	0.0527	NaN		
Canarias	0.2517	0.1567	0.3082	0.2639		167.3
s.e. (Canarias)	0.2644	0.0113	0.1168	0.0778		
Cantabria	0.4252	0.1508	0.685	0.1658		168.7
s.e. (Cantabria)	0.3539	0.0144	0.0938	0.0327		
Castilla La Mancha	0.5518	0.2229	0.5828	0.4961		272.5
s.e. (Castilla La Mancha)	0.1219	0.0054	0.1058	0.0213		
Castilla Leon	-0.0568	0.2477	0.6069	0.5457		245.5
s.e. (Castilla Leon)	0.1649	0.0072	0.116	0.0248		
Catalunya	1.3318	0.2293	0.5496	0.4246		537.4
s.e. (Catalunya)	0.0771	0.0032	0.1002	0.016		
Extremadura	-0.0615	0.1774	0.4034	0.0526		197.2
s.e. (Extremadura)	0.2678	0.0114	0.1025	0.0318		
Galicia	-0.3754	0.2644	0.8832	0.3555		249.6
s.e. (Galicia)	0.1654	0.0084	0.0669	0.0382		
La Rioja	1.9887	0.1006	0.5045	0.3555		261.3
s.e. (La Rioja)	0.1462	0.0067	0.0998	0.0328		
Madrid	3.7197	0.147	0.5185	0.4043		1299.4
s.e. (Madrid)	0.0389	0.0017	0.0962	0.0079		
Murcia	-1.3432	0.22	0.3552	0.5103		141
s.e. (Murcia)	0.3226	0.0148	0.1603	0.0677		
Navarra	0.0996	0.2134	0.714	0.4291		262.2
s.e. (Navarra)	0.214	0.01	0.101	0.0326		
Pais Vasco	1.9849	0.1824	0.8298	0.4462		381.3
s.e. (Pais Vasco)	0.1128	0.0054	0.0773	0.023		
Valencia	4.5479	0.0396	0.6296	0.0557		444.4
s.e. (Valencia)	0.1554	0.0067	0.0929	0.007		

Table 2: Estimates of under-reporting parameters by region in Spain

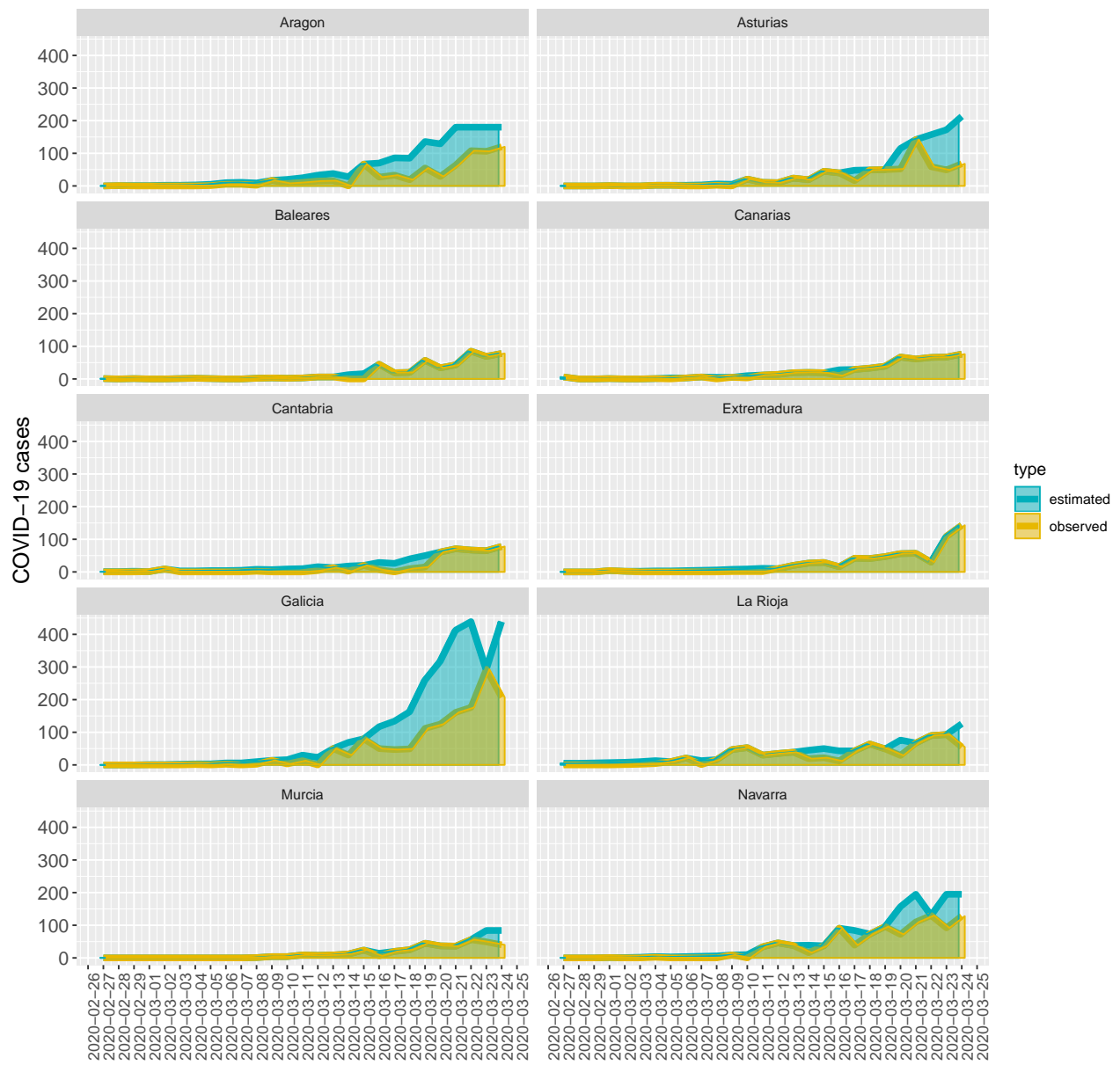


Figure 2 (a): Observed and truly daily cases

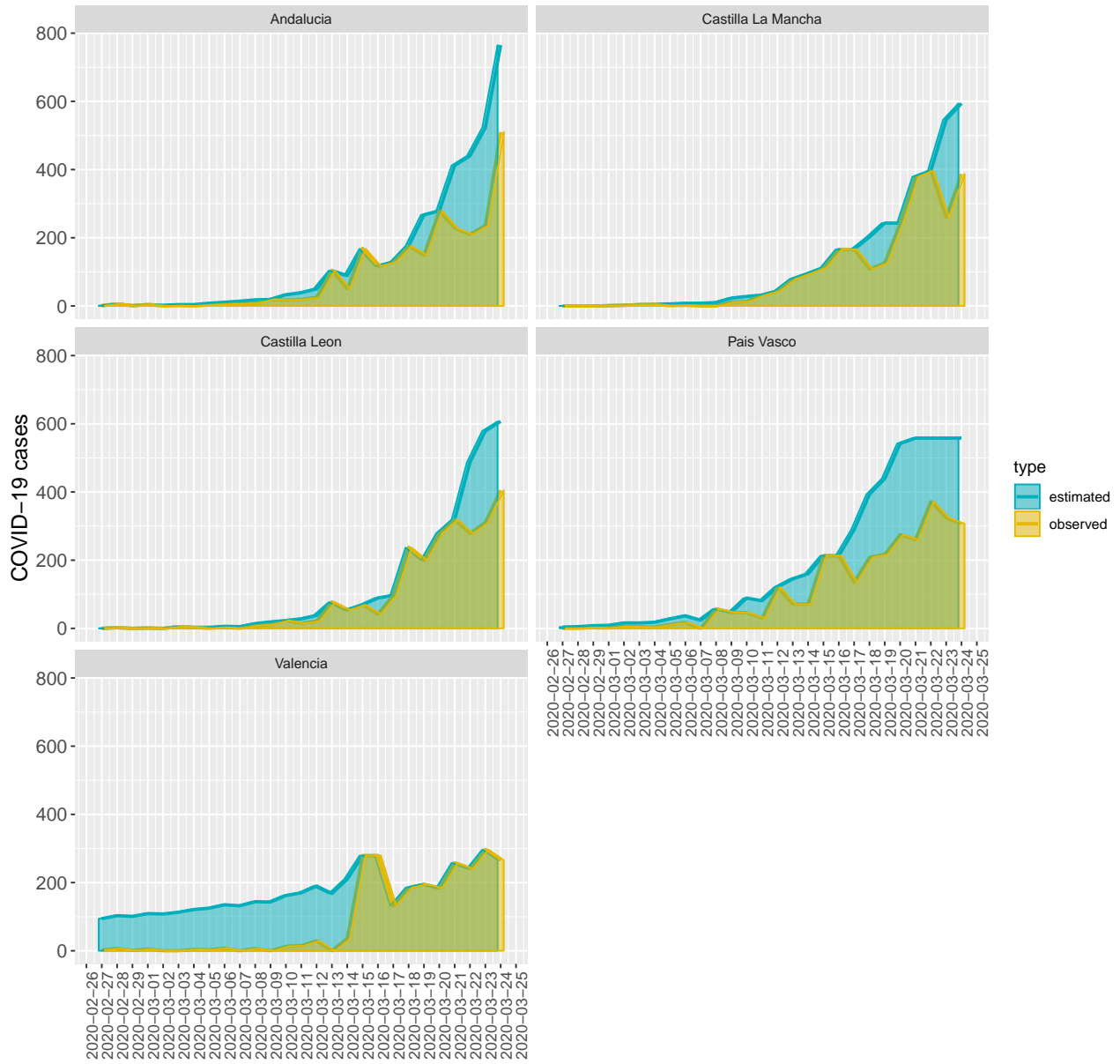


Figure 2 (b): Observed and truly daily cases

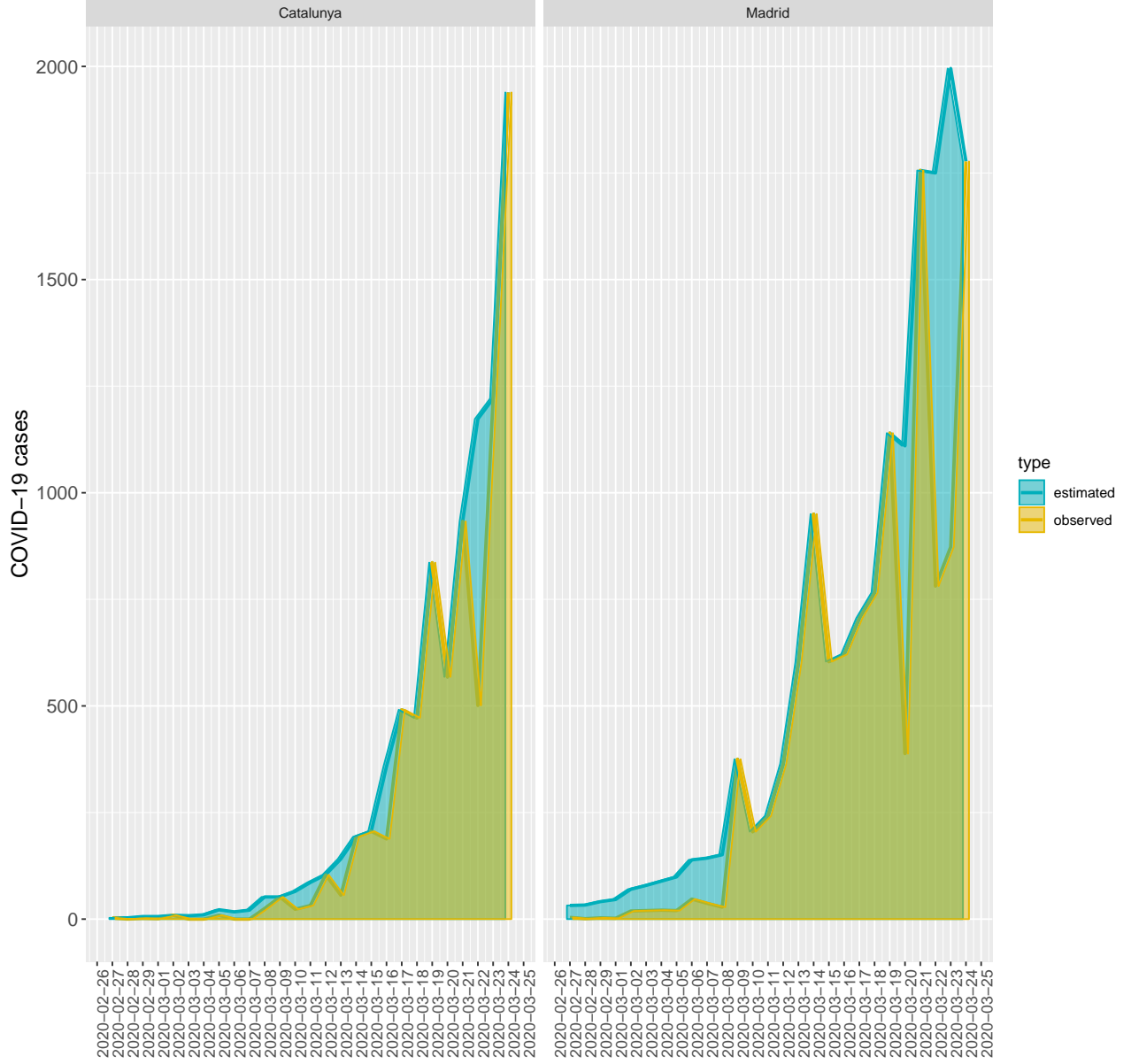


Figure 2 (c): Observed and estimated daily cases

Using the Viterbi algorithm, the model also enables reconstructing the most likely sequence of real COVID-19 cases throughout the study. This allows us to have an estimated time series of truly daily cases and evaluate the impact of under-reporting over measures such as the basic reproduction number. Figure 2 shows the observed and reconstructed series over time by region.

Table 3 shows the percentages of means counts that are not covered by the official registers. Thus, the highest the rate, the lower is the coverage, and therefore the severe is the impact of the under-reporting.

	observed mean	true mean	% not covered
Andalucia	91.52	136.41	32.91
Aragon	26.26	55.52	52.70
Asturias	24.52	42.11	41.78
Baleares	17.70	18.78	5.72
Canarias	20.63	21.85	5.59
Cantabria	15.74	23.00	31.56
Castilla La Mancha	97.44	125.63	22.44
Castilla Leon	91.11	120.04	24.10
Catalunya	291.26	333.15	12.57
Extremadura	23.56	25.33	7.02
Galicia	52.41	107.04	51.04
La Rioja	29.70	39.41	24.62
Madrid	457.48	588.33	22.24
Murcia	14.26	17.48	18.43
Navarra	37.56	53.89	30.31
Pais Vasco	111.56	192.07	41.92
Valencia	90.59	172.00	47.33

Table 3: Estimate mean of non-coverage of cases of COVID-19 in Spain

CCAA	estimated	observed	lethality-estimated	lethality-observed	mortality
Andalucia	3683.00	2471.00	2.36	3.52	1.03
Aragon	1499.00	709.00	2.47	5.22	2.80
Asturias	1137.00	662.00	1.93	3.32	2.15
Baleares	507.00	478.00	1.97	2.09	0.87
Canarias	590.00	557.00	2.71	2.87	0.74
Cantabria	621.00	425.00	1.45	2.12	1.55
Castilla La Mancha	3392.00	2631.00	6.37	8.21	10.63
Castilla Leon	3241.00	2460.00	3.83	5.04	5.17
Catalunya	8995.00	7864.00	3.14	3.59	3.67
Extremadura	684.00	636.00	3.80	4.09	2.44
Galicia	2890.00	1415.00	0.69	1.41	0.74
La Rioja	1064.00	802.00	2.82	3.74	9.47
Madrid	15885.00	12352.00	9.66	12.43	23.04
Murcia	472.00	385.00	0.64	0.78	0.20
Navarra	1455.00	1014.00	2.13	3.06	4.74
Pais Vasco	5186.00	3012.00	2.56	4.42	6.02
Valencia	4644.00	2446.00	2.48	4.70	2.30

Table 4: Estimated mortality and lethality rates