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 juste announced a new protocol with rapid tests to be implementend 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.

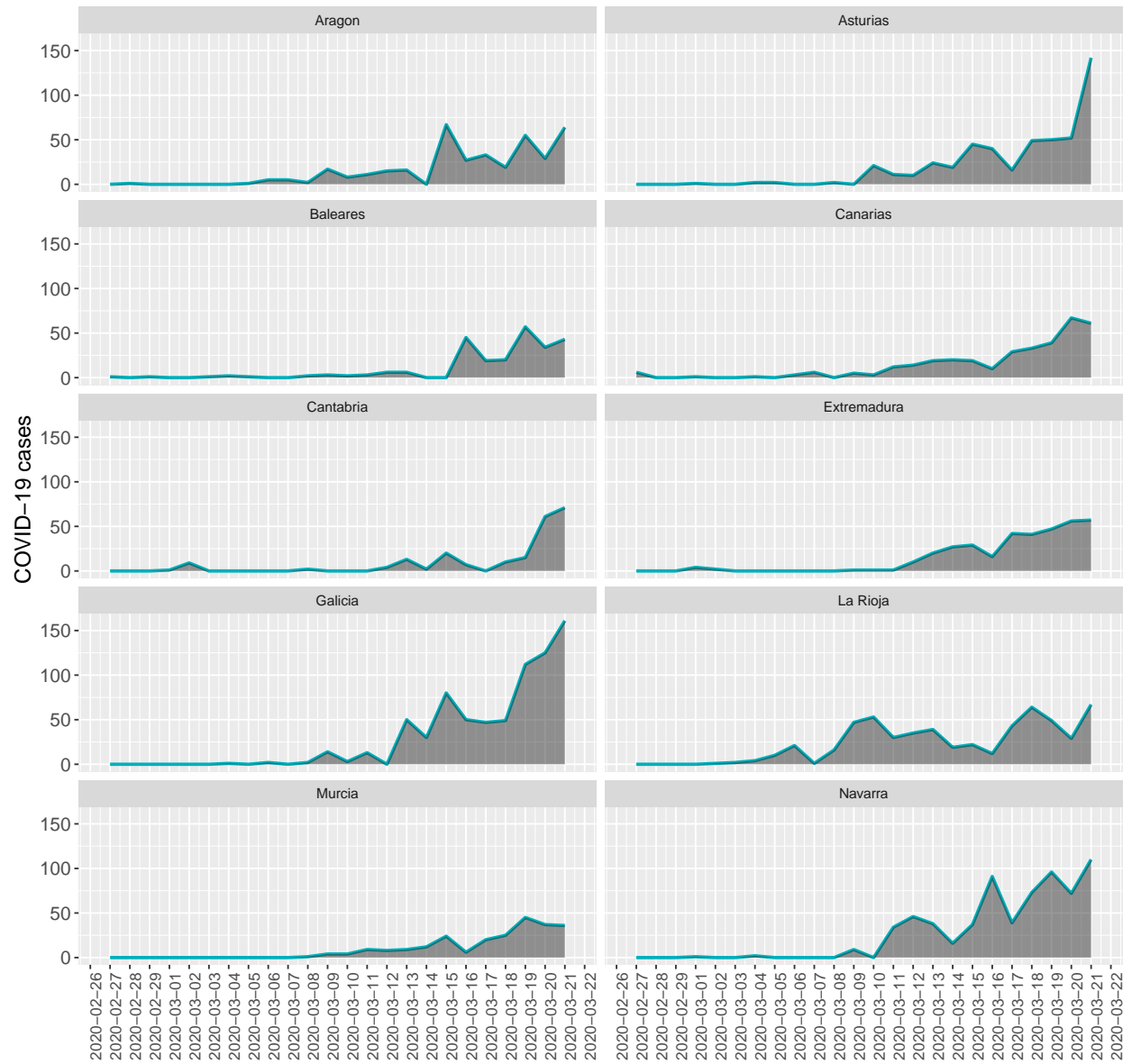Notice that this analysis con be easily reproduced for other countries.

Figure 1 (a): Daily COVID−19 cases from 27−02−2020 to 21−03−2020
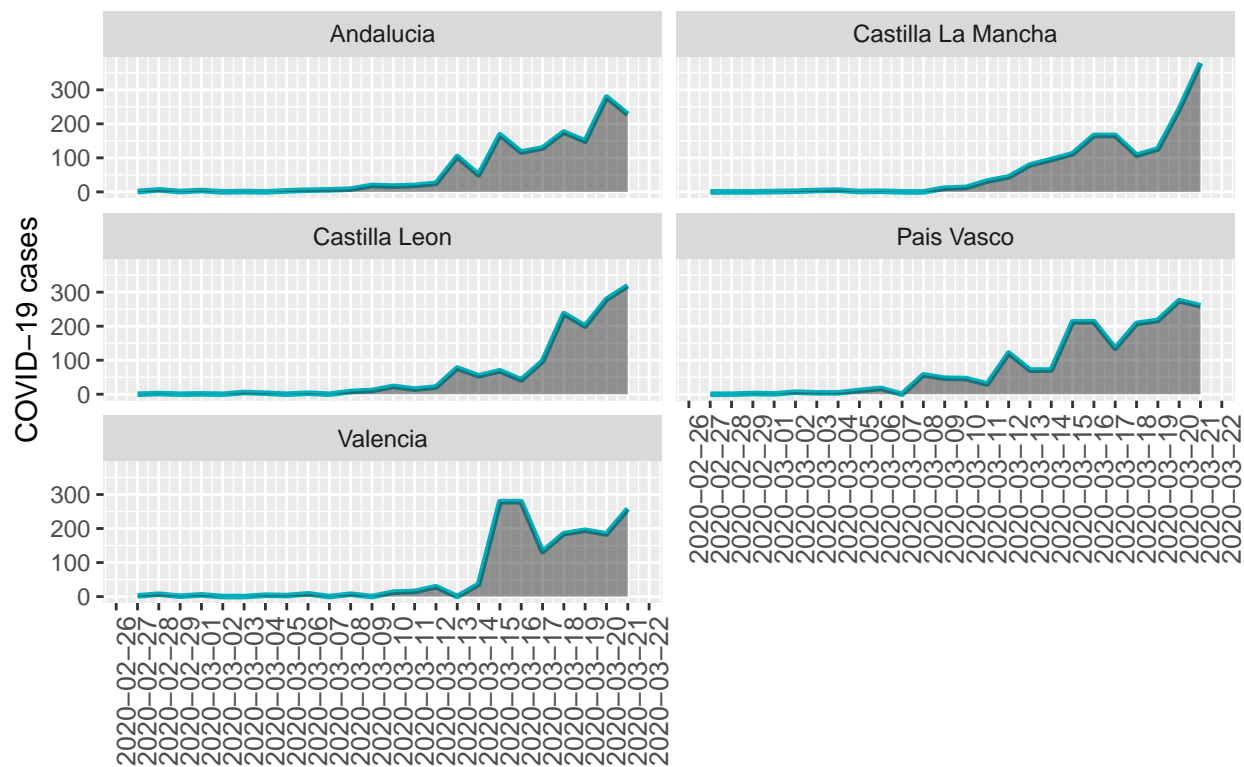
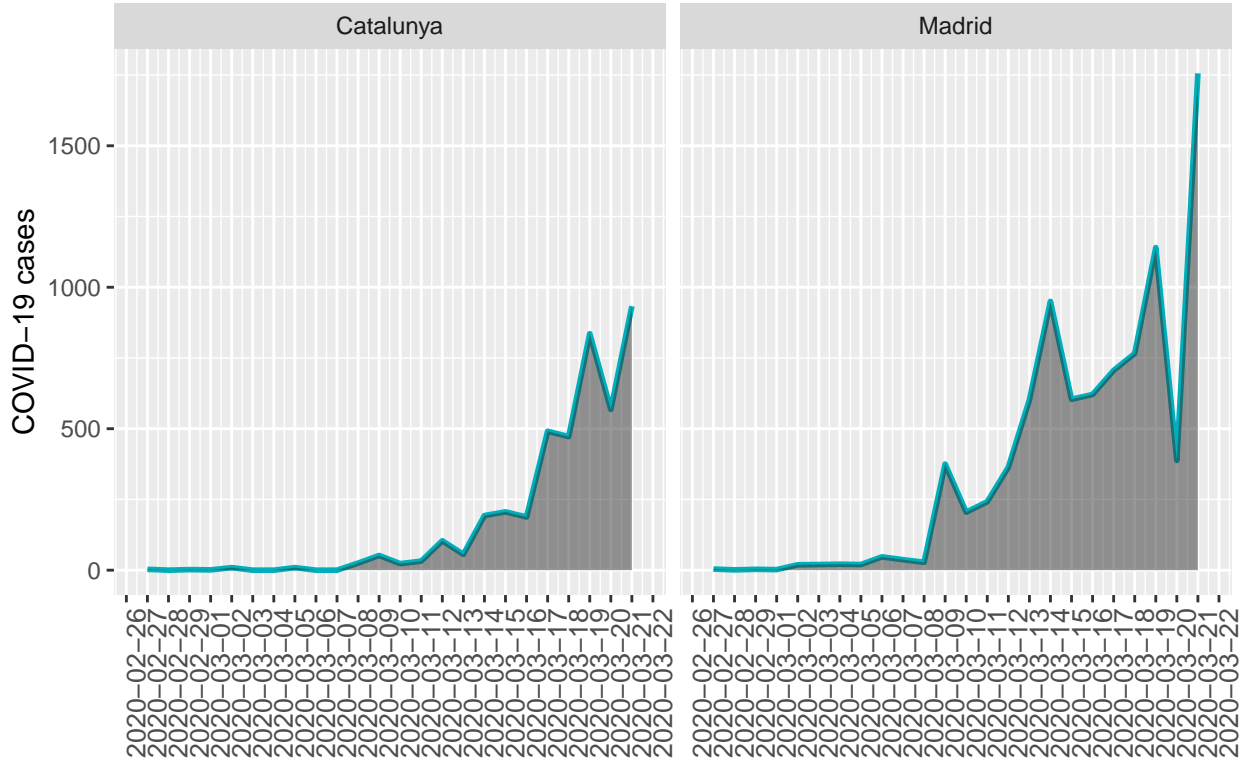Figure 1 (b): Daily COVID−19 cases from 27−02−2020 to 21−03−2020

Figure 1 (c): Daily COVID−19 cases from 27−02−2020 to 21−03−2020

	minimum	mean	median	maximum	standard deviation	dispersion index
Andalucia	0.00	63.12	18.00	279.00	83.68	110.93
Aragon	0.00	15.62	6.50	67.00	20.62	27.22
Asturias	0.00	20.25	6.00	142.00	31.91	50.29
Baleares	0.00	10.25	2.00	57.00	16.97	28.10
Canarias	0.00	14.50	6.00	67.00	18.99	24.87
Cantabria	0.00	8.96	0.50	71.00	18.52	38.27
Castilla La Mancha	0.00	66.21	12.00	379.00	95.84	138.72
Castilla Leon	0.00	61.08	13.00	319.00	96.10	151.20
Catalunya	0.00	175.12	28.50	933.00	276.14	435.43
Extremadura	0.00	14.75	1.50	57.00	19.96	27.02
Galicia	0.00	30.79	2.50	161.00	45.98	68.66
La Rioja	0.00	23.50	20.00	67.00	21.76	20.16
Madrid	1.00	371.71	223.50	1756.00	451.95	549.51
Murcia	0.00	10.00	4.00	45.00	13.72	18.83
Navarra	0.00	27.67	5.50	110.00	35.93	46.65
Pais Vasco	0.00	83.71	46.50	275.00	95.11	108.06
Valencia	0.00	68.42	7.50	279.00	101.34	150.09

Table 1: Summary of the daily COVID-19 cases from 27-02-20 to 21-03-2020 by region in Spain

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 growths 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 and observed processes.

	α	β_0	β_1	ω	q	AIC
Andalucia	0.673	0.2204	0.6747	0.5483	243.1
s.e. (Andalucia)	0.1529	0.0073	0.1064	0.0301	
Aragon	0.0145	0.1833	0.425	0.325	191.7
s.e. (Aragon)	0.2757	0.0135	0.1292	0.0508	
Asturias	-0.8127	0.2429	0.64	0.4796	151.2
s.e. (Asturias)	0.3535	0.0154	0.158	0.0454	
Baleares	-1.3348	0.253	0.8324	0.3497	133.9
s.e. (Baleares)	0.4577	0.0211	0.1115	0.0459	
Canarias	-0.2329	0.1836	0.2899	0.2918	139.3
s.e. (Canarias)	0.3236	0.0154	0.1326	0.0893	
Cantabria	-0.2446	0.1884	0.7576	0.1688	131.4
s.e. (Cantabria)	0.4443	0.0199	0.0967	0.0326	
Castilla La Mancha	0.4181	0.2316	0.5536	0.4662	227.3
s.e. (Castilla La Mancha)	0.1572	0.0074	0.1168	0.0322	
Castilla Leon	-0.7038	0.2974	0.7997	0.542	184.3
s.e. (Castilla Leon)	0.1893	0.0091	0.1067	0.0354	
Catalunya	-0.3254	0.3197	0.502	0.552	274.2
s.e. (Catalunya)	0.1155	0.0057	0.1148	0.0187	
Extremadura	0.3053	0.1605	0.4788	0.0432	124.5
s.e. (Extremadura)	0.349	0.0167	0.1073	0.0264	
Galicia	-0.3338	0.23	0.558	0.4417	203.2
s.e. (Galicia)	0.2814	0.0126	0.1354	0.0596	
La Rioja	2.1782	0.0911	0.5311	0.2922	228.7
s.e. (La Rioja)	0.1797	0.0092	0.1059	0.0339	
Madrid	3.61	0.1578	0.4999	0.2293	820.5
s.e. (Madrid)	0.05434	0.0027	0.1021	0.01	
Murcia	-1.4065	0.2226	0.2018	0.2974	112.3
s.e. (Murcia)	0.4173	0.0196	0.1607	0.1403	
Navarra	-0.4287	0.2606	0.7956	0.3586	179.7
s.e. (Navarra)	0.2158	0.0101	0.0932	0.0207	
Pais Vasco	1.3899	0.2166	0.7748	0.4421	246.1
s.e. (Pais Vasco)	0.1545	0.008	0.0926	0.0235	
Valencia	4.6706	0.0335	0.7084	0.0523	410.2
s.e. (Valencia)	0.2236	0.0105	0.0927	0.0079	

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

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.

Figure 2 (a): Observed and truly daily cases
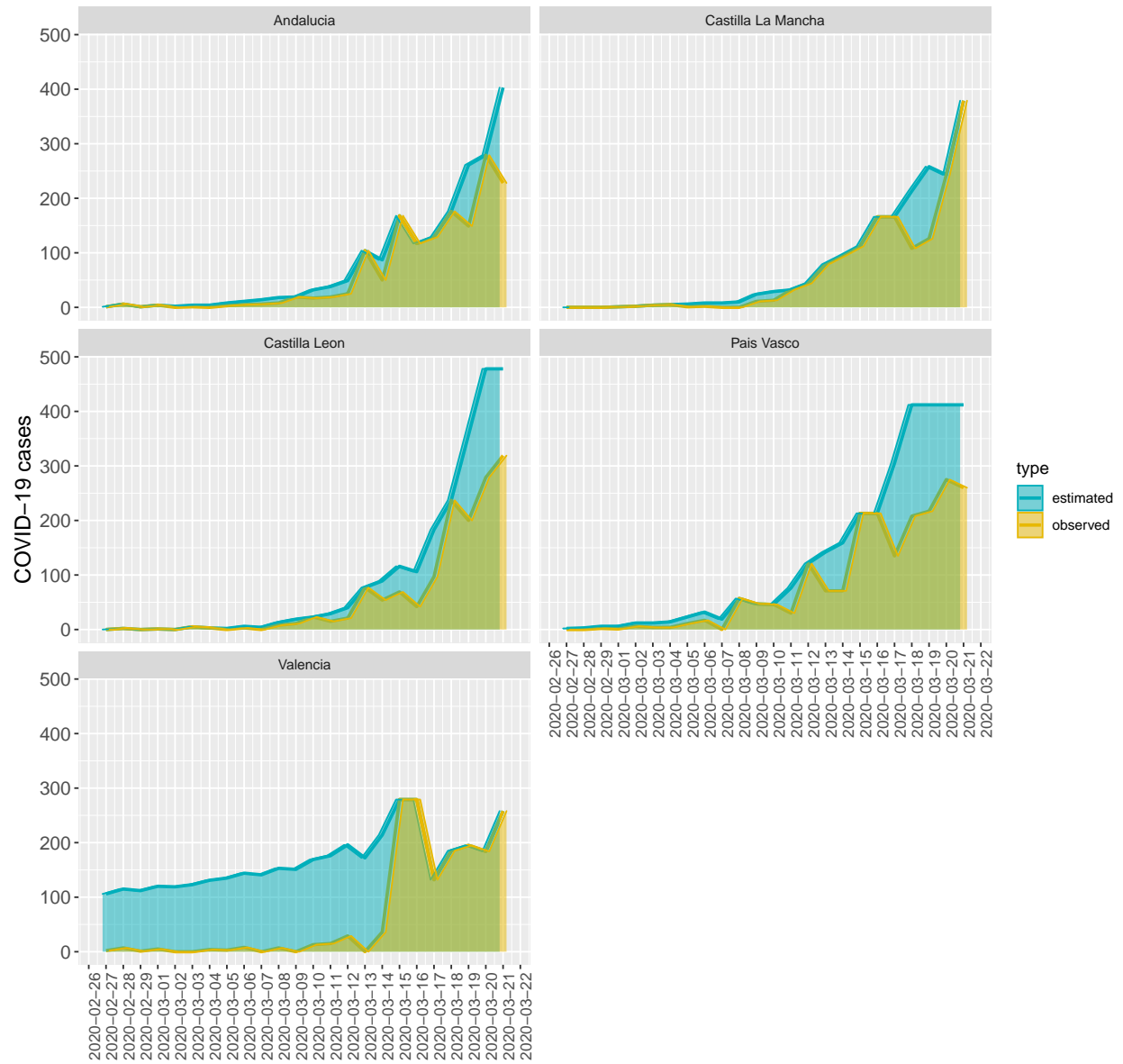
Figure 2 (b): Observed and truly daily cases
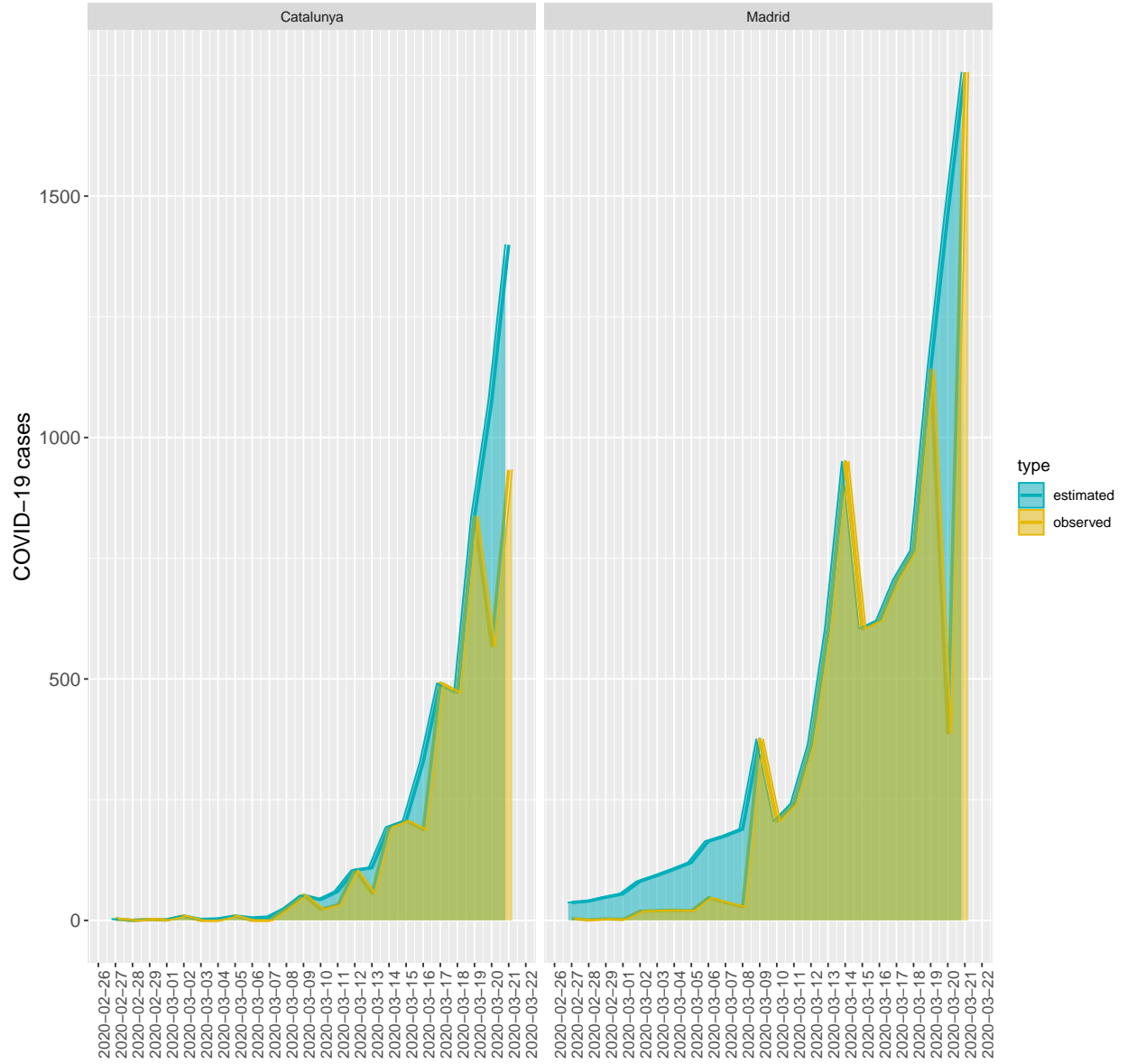
Figure 2 (c): Observed and estimated daily cases

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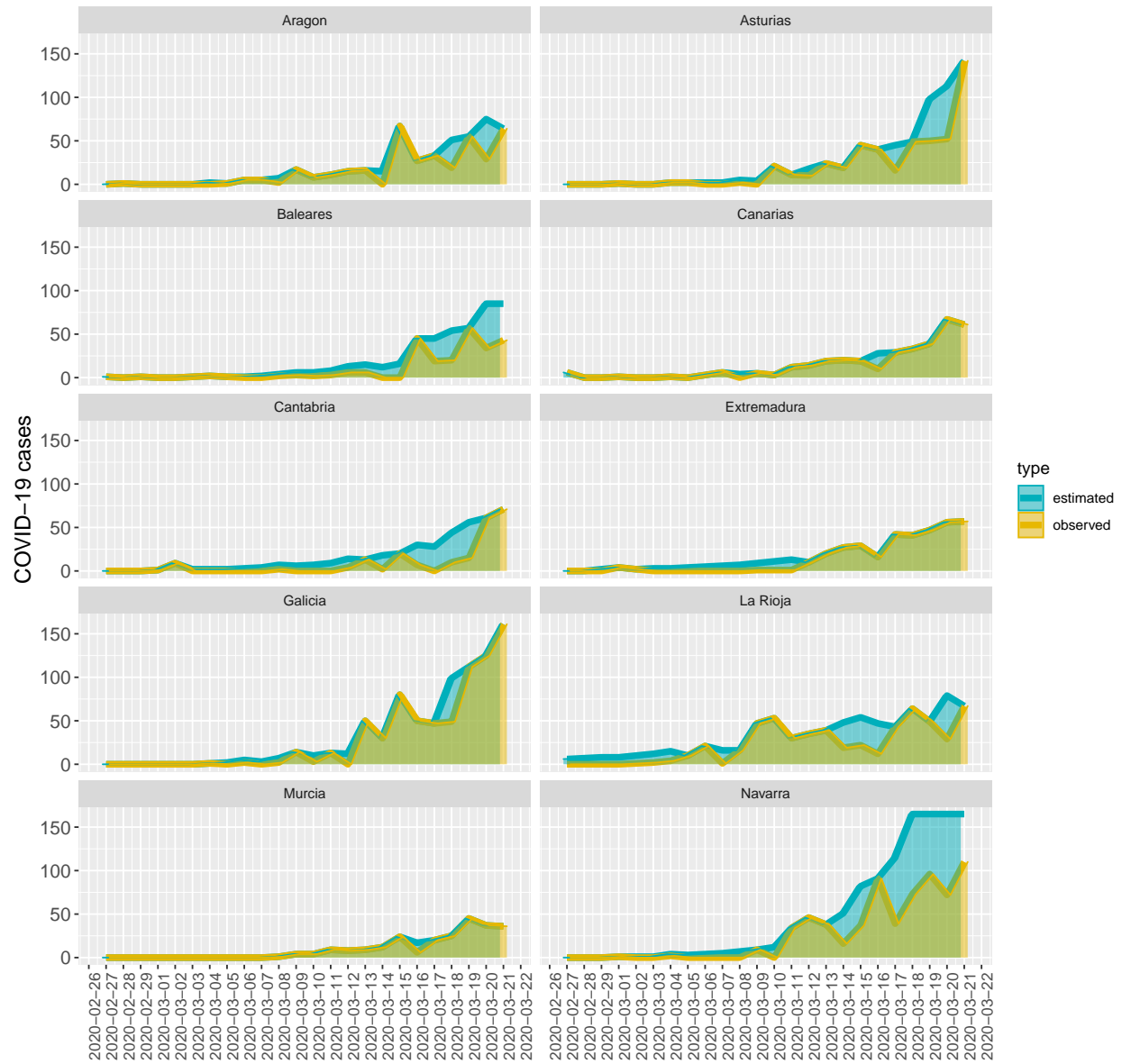
Table 3 shows the percentages of means counts that are not covered by the official registers. Thus, the highest the rate, the lower is the coverage, and therefore the severe is the impact of the under-reporting.

	observed mean	true mean	% not covered
Andalucia	63.12	80.62	21.71
Aragon	15.62	19.79	21.05
Asturias	20.25	26.79	24.42
Baleares	10.25	19.17	46.52
Canarias	14.50	15.42	5.95
Cantabria	8.96	16.96	47.17
Castilla La Mancha	66.21	78.50	15.66
Castilla Leon	61.08	94.50	35.36
Catalunya	175.12	226.33	22.63
Extremadura	14.75	17.25	14.49
Galicia	30.79	34.21	9.99
La Rioja	23.50	32.67	28.06
Madrid	371.71	454.17	18.16
Murcia	10.00	10.46	4.38
Navarra	27.67	48.50	42.96
Pais Vasco	83.71	131.50	36.34
Valencia	68.42	166.21	58.84

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

CCAA	estimated	observed	lethality-estimated	lethality-observed	mortality
Andalucia	1935.00	1515.00	2.07	2.64	0.48
Aragon	475.00	375.00	4.63	5.87	1.67
Asturias	643.00	486.00	0.78	1.03	0.49
Baleares	460.00	246.00	0.87	1.63	0.35
Canarias	370.00	348.00	1.89	2.01	0.33
Cantabria	407.00	215.00	0.49	0.93	0.34
Castilla La Mancha	1884.00	1589.00	4.46	5.29	4.13
Castilla Leon	2268.00	1466.00	2.43	3.75	2.29
Catalunya	5432.00	4203.00	2.25	2.90	1.59
Extremadura	414.00	354.00	2.90	3.39	1.12
Galicia	821.00	739.00	1.10	1.22	0.33
La Rioja	784.00	564.00	1.91	2.66	4.73
Madrid	10900.00	8921.00	7.38	9.01	12.07
Murcia	251.00	240.00	0.40	0.42	0.07
Navarra	1164.00	664.00	0.77	1.36	1.38
Pais Vasco	3156.00	2009.00	2.69	4.23	3.85
Valencia	3989.00	1642.00	1.25	3.05	1.00

Table 4: Estimated mortality and lethality rates