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.
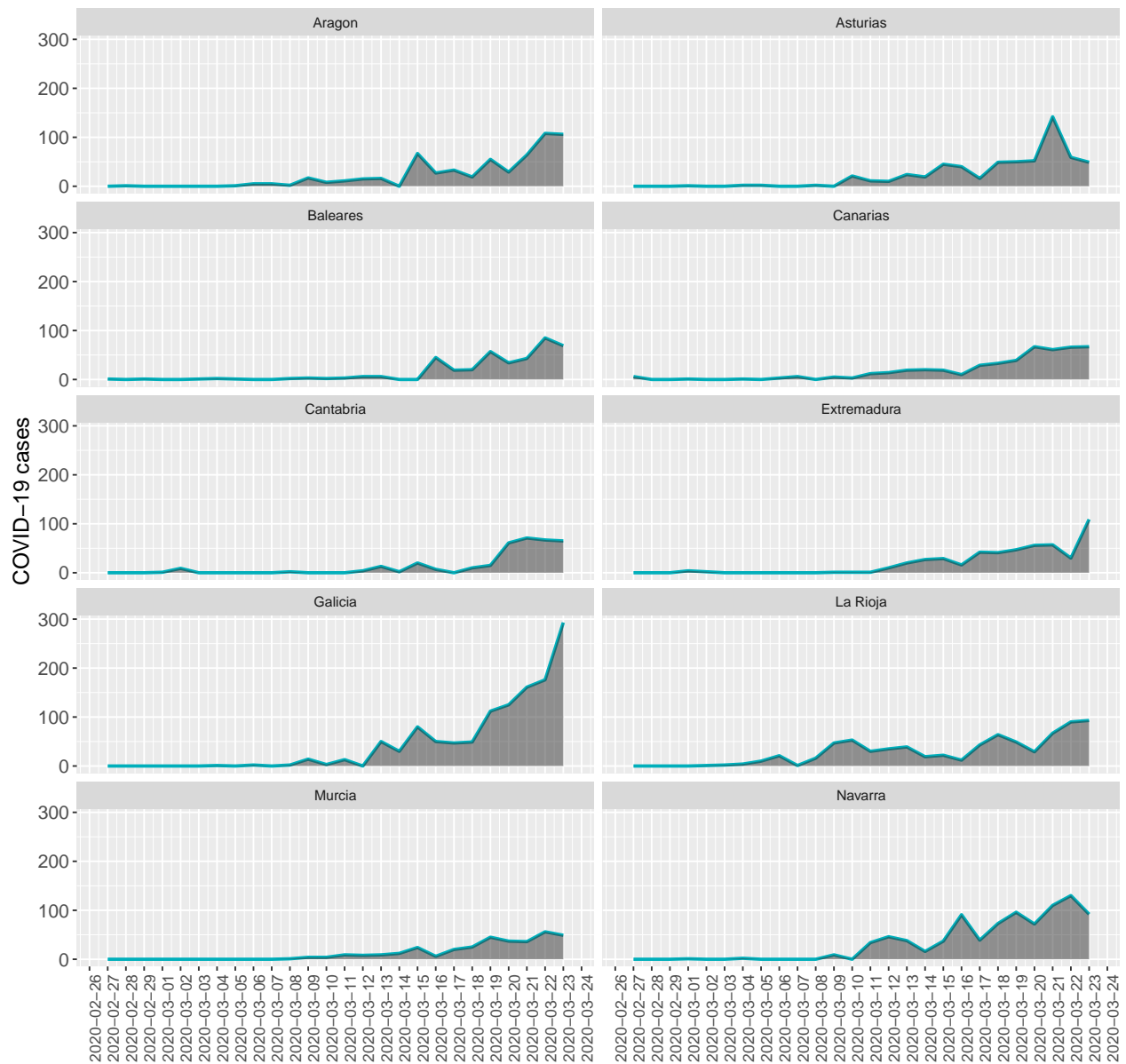
Figure 1 (a): Daily COVID−19 cases from 27−02−2020 to 23−03−2020
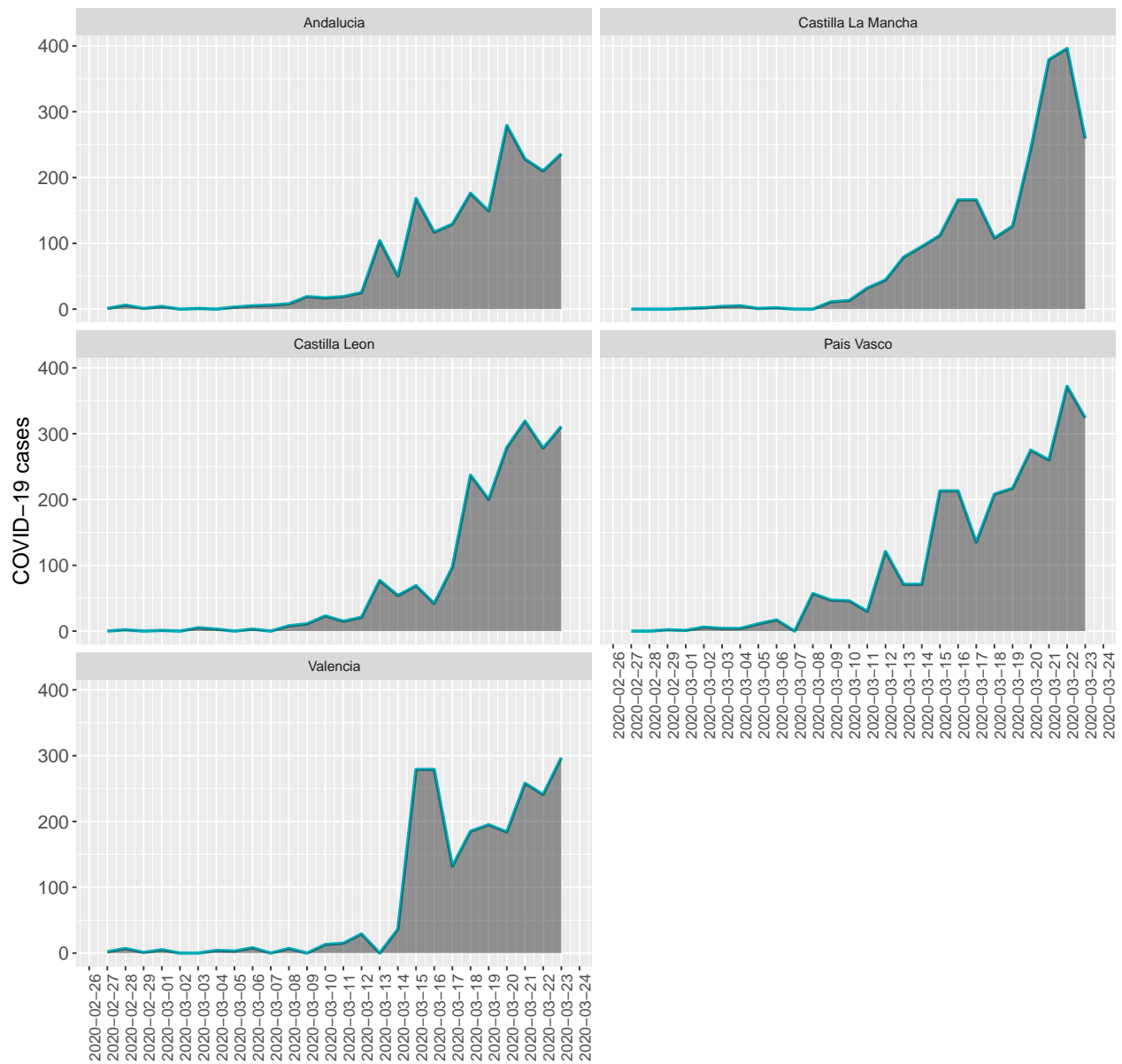
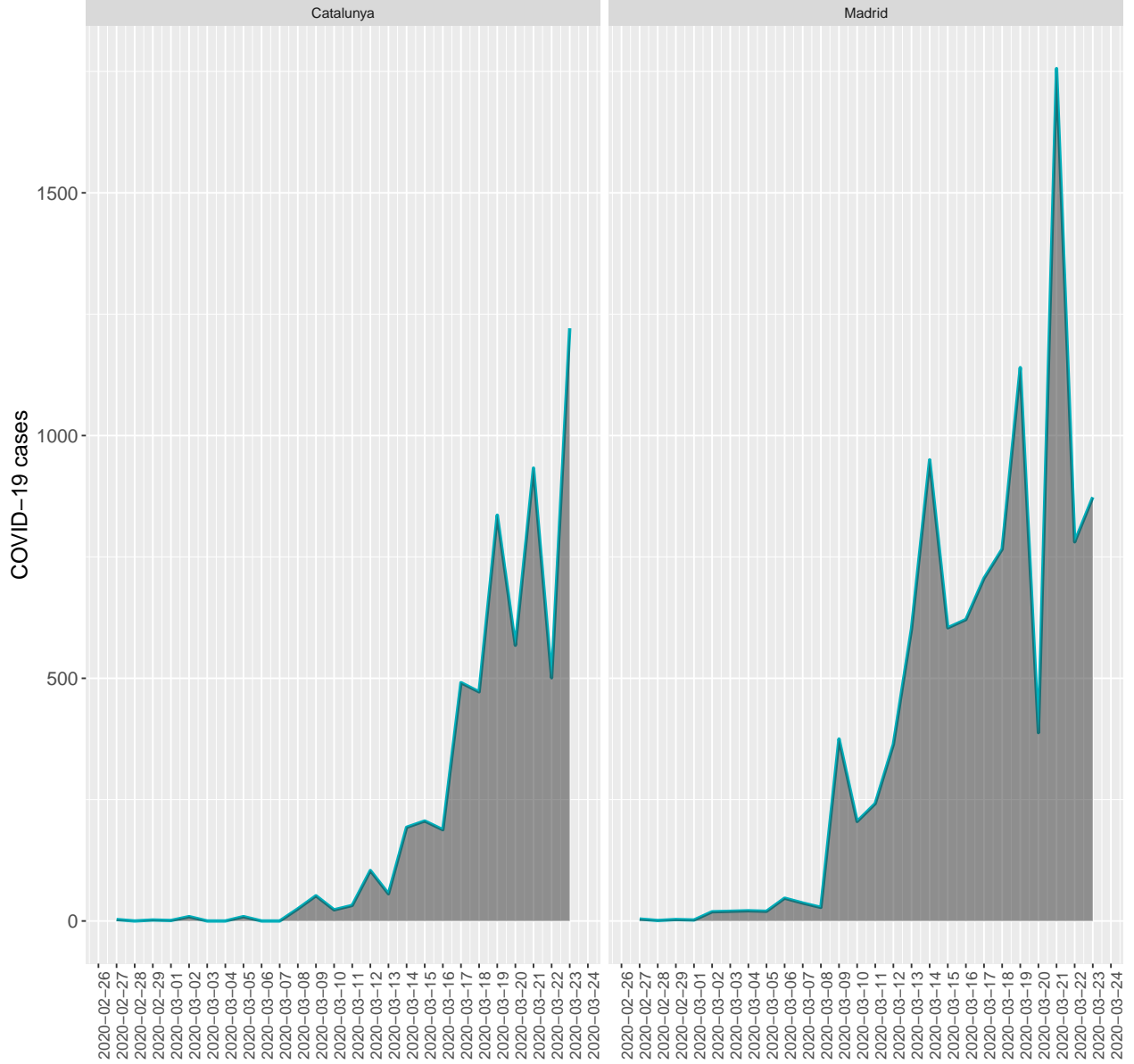Figure 1 (b): Daily COVID−19 cases from 27−02−2020 to 23−03−2020

Figure 1 (c): Daily COVID–19 cases from 27–02–2020 to 23–03–2020

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 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

4

	minimum	mean	median	maximum	standard deviation	dispersion index
Andalucia	0.00	75.42	19.00	279.00	91.34	110.62
Aragon	0.00	22.65	9.50	108.00	31.75	44.49
Asturias	0.00	22.85	10.50	142.00	31.98	44.77
Baleares	0.00	15.38	2.00	85.00	24.48	38.94
Canarias	0.00	18.50	8.00	67.00	23.05	28.73
Cantabria	0.00	13.35	1.50	71.00	23.57	41.64
Castilla La Mancha	0.00	86.31	22.50	396.00	117.76	160.67
Castilla Leon	0.00	79.04	18.00	319.00	111.99	158.69
Catalunya	0.00	227.88	42.00	1221.00	339.50	505.79
Extremadura	0.00	18.96	3.00	109.00	26.70	37.59
Galicia	0.00	46.46	8.00	293.00	72.69	113.71
La Rioja	0.00	28.73	21.50	93.00	27.88	27.06
Madrid	1.00	406.73	303.00	1756.00	450.99	500.07
Murcia	0.00	13.27	5.00	56.00	17.54	23.18
Navarra	0.00	34.08	12.50	130.00	41.58	50.74
Pais Vasco	0.00	104.04	52.00	372.00	116.30	130.01
Valencia	0.00	83.85	10.50	297.00	111.72	148.86

Table 1: Summary of the daily COVID-19 cases from 27-02-20 to 23-03-2020 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	0.9905	0.2049	0.7213	0.4906	274.1
s.e. (Andalucia)	0.1229	0.0058	0.0934	0.023	
Aragon	0.0909	0.2181	0.909	0.3293	197.8
s.e. (Aragon)	0.2196	0.0091	0.0619	0.0215	
Asturias	-0.6889	0.229	0.5055	0.3546	187
s.e. (Asturias)	0.2955	0.0138	0.1418	0.0371	
Baleares	-0.7582	0.2065	0.7462	0.5023	183.8
s.e. (Baleares)	0.3918	0.0158	0.1206	0.0609	
Canarias	0.0867	0.1655	0.2999	0.2752	157.6
s.e. (Canarias)	0.2822	0.0125	0.1219	0.0827	
Cantabria	0.2432	0.1602	0.7077	0.1695	157.6
s.e. (Cantabria)	0.383	0.016	0.0948	0.0334	
Castilla La Mancha	0.6549	0.2182	0.5749	0.4678	258
s.e. (Castilla La Mancha)	0.1309	0.0058	0.1073	0.0235	
Castilla Leon	-0.0959	0.2495	0.5799	0.5504	236.3
s.e. (Castilla Leon)	0.1764	0.0077	0.1225	0.0264	
Catalunya	1.4342	0.224	0.5824	0.4279	518.3
s.e. (Catalunya)	0.0847	0.0037	0.1013	0.01613	
Extremadura	0.4152	0.1533	0.4428	0.0417	176.9
s.e. (Extremadura)	0.2842	0.0127	0.1018	0.0252	
Galicia	-0.931	0.2947	0.8647	0.3389	182
s.e. (Galicia)	0.1804	0.0075	0.0736	0.0174	
La Rioja	2.1611	0.0921	0.4868	0.2915	244.4
s.e. (La Rioja)	0.1607	0.0075	0.1016	0.0338	
Madrid	3.5683	0.1557	0.5385	0.4079	1186.8
s.e. (Madrid)	0.0421	0.0019	0.0978	0.0076	
Murcia	-0.937	0.1968	0.233	0.265	132.4
s.e. (Murcia)	0.3768	0.0164	0.1548	0.1298	
Navarra	-0.0791	0.2224	0.6861	0.4373	249.3
s.e. (Navarra)	0.2282	0.0109	0.1089	0.0339	
Pais Vasco	1.6895	0.1993	0.8068	0.4407	302.3
s.e. (Pais Vasco)	0.0919	0.0049	0.078	0.023	
Valencia	4.4958	0.0421	0.6538	0.0572	434.5
s.e. (Valencia)	0.1745	0.0078	0.0933	0.0075	

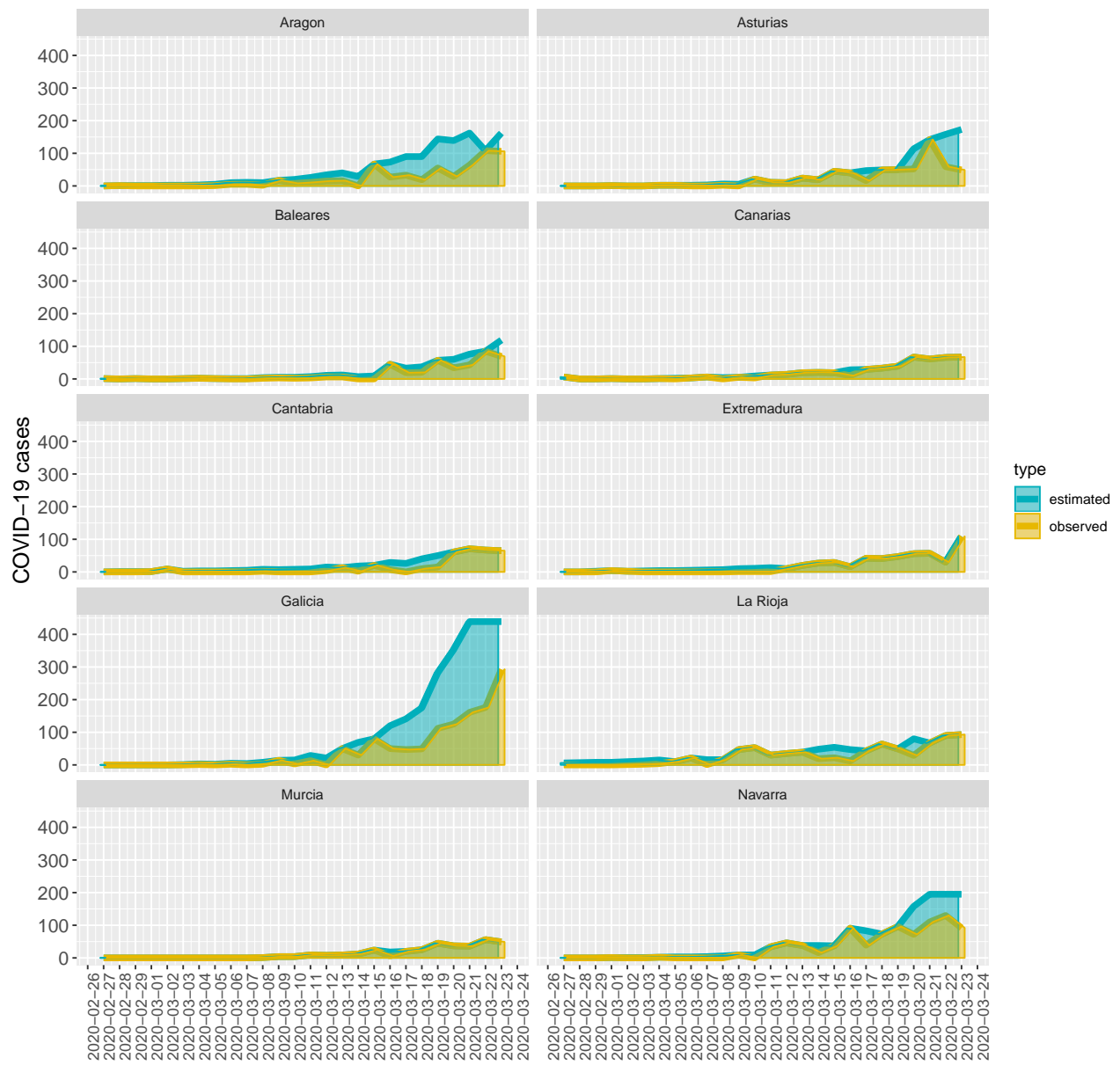Table 2: Estimates of under-reporting parameters by region in Spain
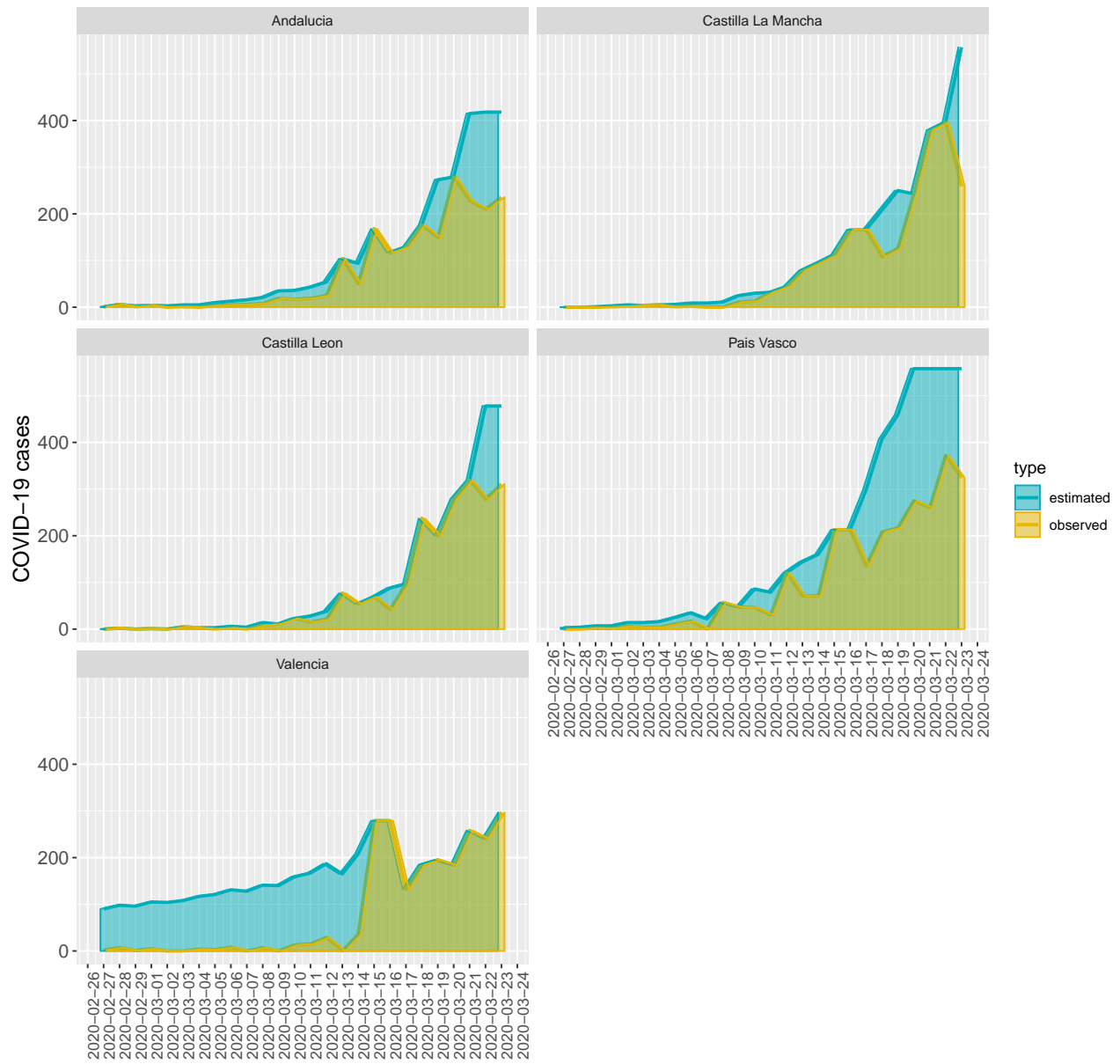
Figure 2 (a): Observed and truly daily cases

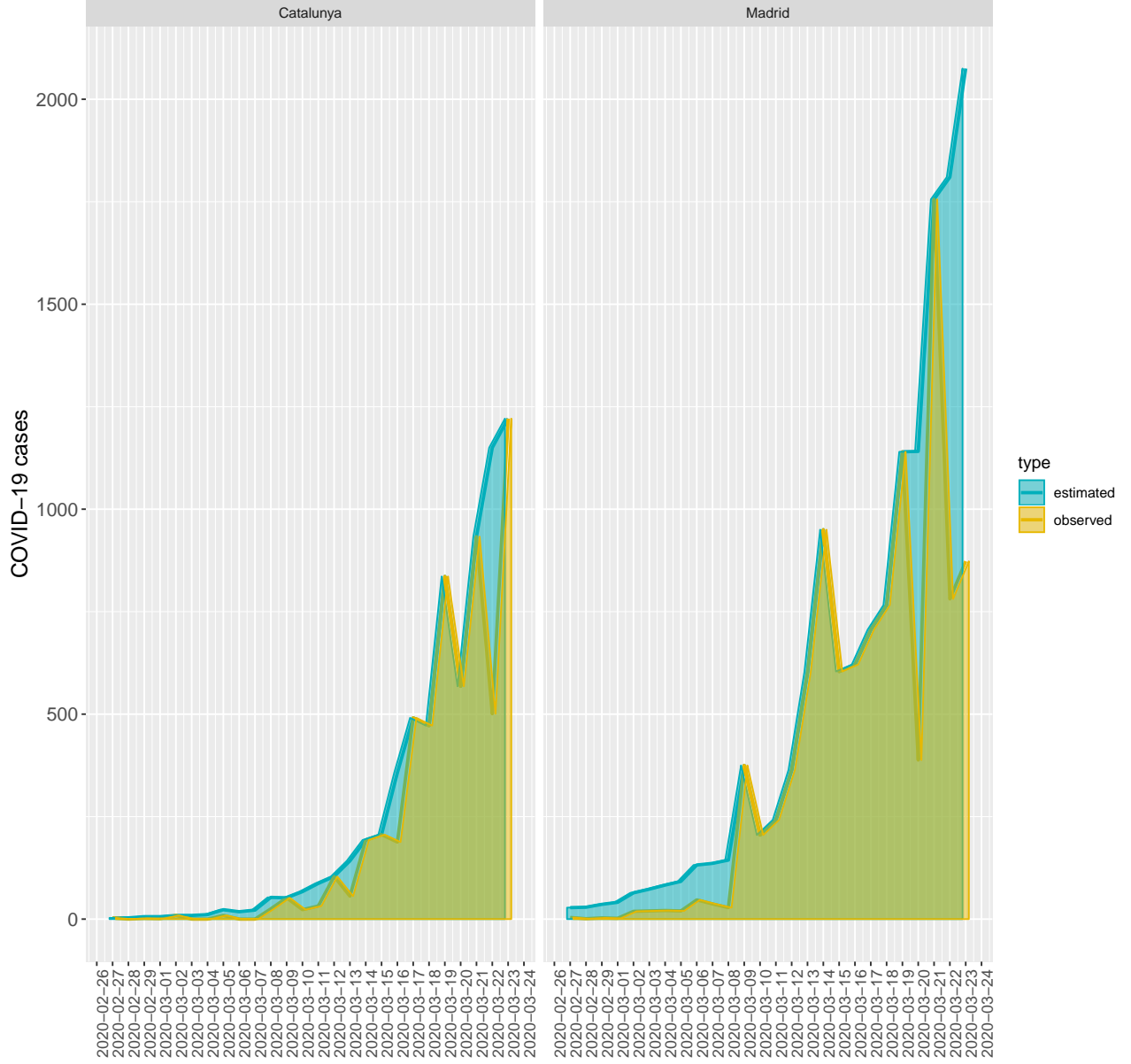Figure 2 (b): Observed and truly daily cases

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	75.42	109.46	31.10
Aragon	22.65	47.96	52.77
Asturias	22.85	35.54	35.71
Baleares	15.38	22.31	31.03
Canarias	18.50	19.65	5.87
Cantabria	13.35	20.62	35.26
Castilla La Mancha	86.31	109.04	20.85
Castilla Leon	79.04	96.69	18.26
Catalunya	227.88	270.85	15.86
Extremadura	18.96	21.35	11.17
Galicia	46.46	103.42	55.08
La Rioja	28.73	37.23	22.83
Madrid	406.73	546.73	25.61
Murcia	13.27	13.73	3.36
Navarra	34.08	50.69	32.78
Pais Vasco	104.04	179.35	41.99
Valencia	83.85	165.96	49.48

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

CCAA	estimated	observed	lethality-estimated	lethality-observed	mortality
Andalucia	2846.00	1961.00	2.04	2.96	0.69
Aragon	1247.00	589.00	2.57	5.43	2.43
Asturias	924.00	594.00	1.30	2.02	1.17
Baleares	580.00	400.00	1.72	2.50	0.87
Canarias	511.00	481.00	2.15	2.29	0.51
Cantabria	536.00	347.00	1.12	1.73	1.03
Castilla La Mancha	2835.00	2244.00	5.11	6.46	7.13
Castilla Leon	2514.00	2055.00	4.06	4.96	4.25
Catalunya	7042.00	5925.00	3.48	4.14	3.19
Extremadura	555.00	493.00	3.24	3.65	1.69
Galicia	2689.00	1208.00	0.67	1.49	0.67
La Rioja	968.00	747.00	2.27	2.95	6.94
Madrid	14215.00	10575.00	8.88	11.94	18.95
Murcia	357.00	345.00	0.56	0.58	0.13
Navarra	1318.00	886.00	1.82	2.71	3.67
Pais Vasco	4663.00	2705.00	2.57	4.44	5.44
Valencia	4315.00	2180.00	2.18	4.31	1.88

Table 4: Estimated mortality and lethality rates