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.

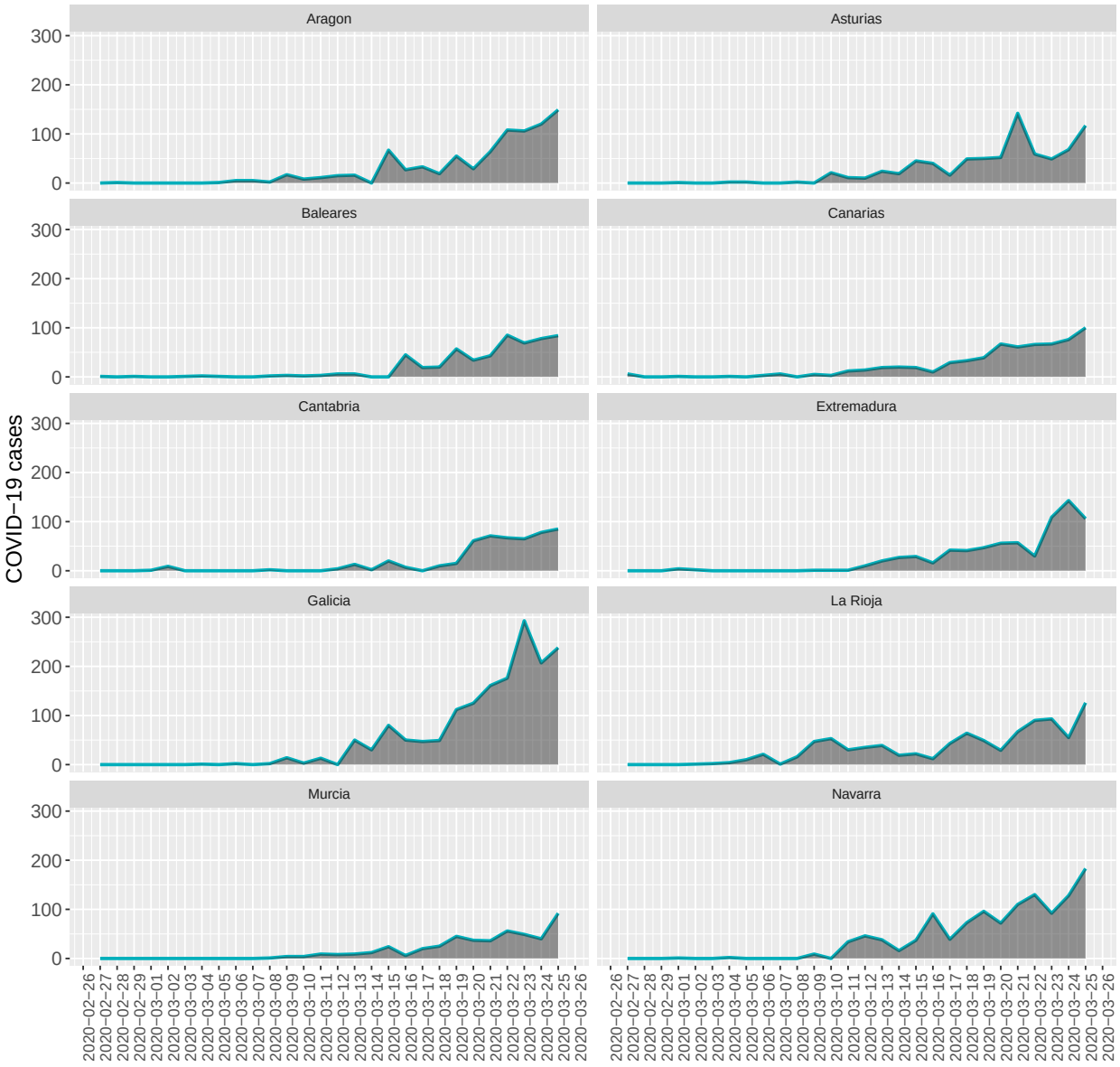

Figure 1 (a): Daily COVID−19 cases from 27−02−2020 to 25−03−2020

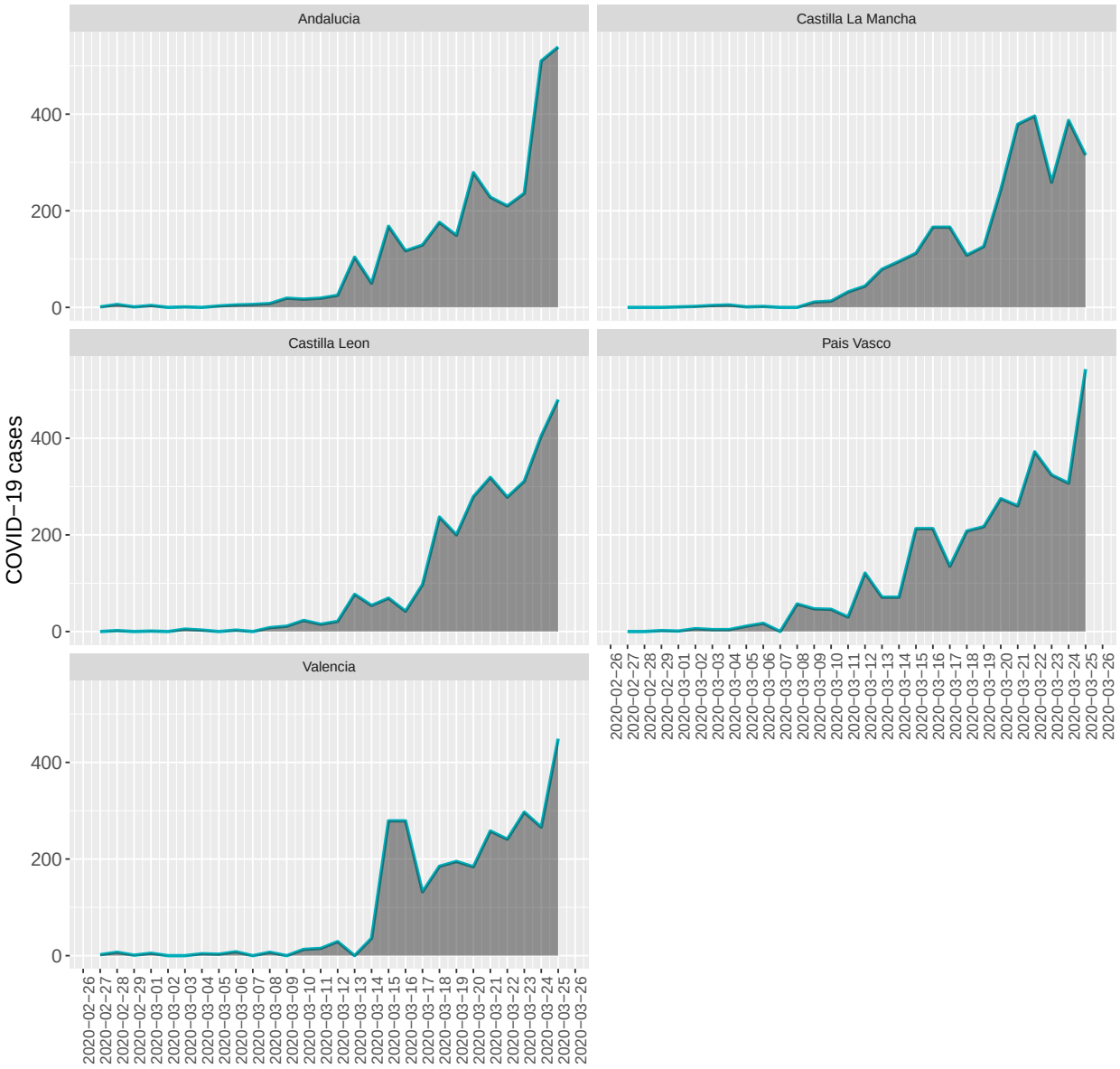

Figure 1 (b): Daily COVID−19 cases from 27−02−2020 to 25−03−2020

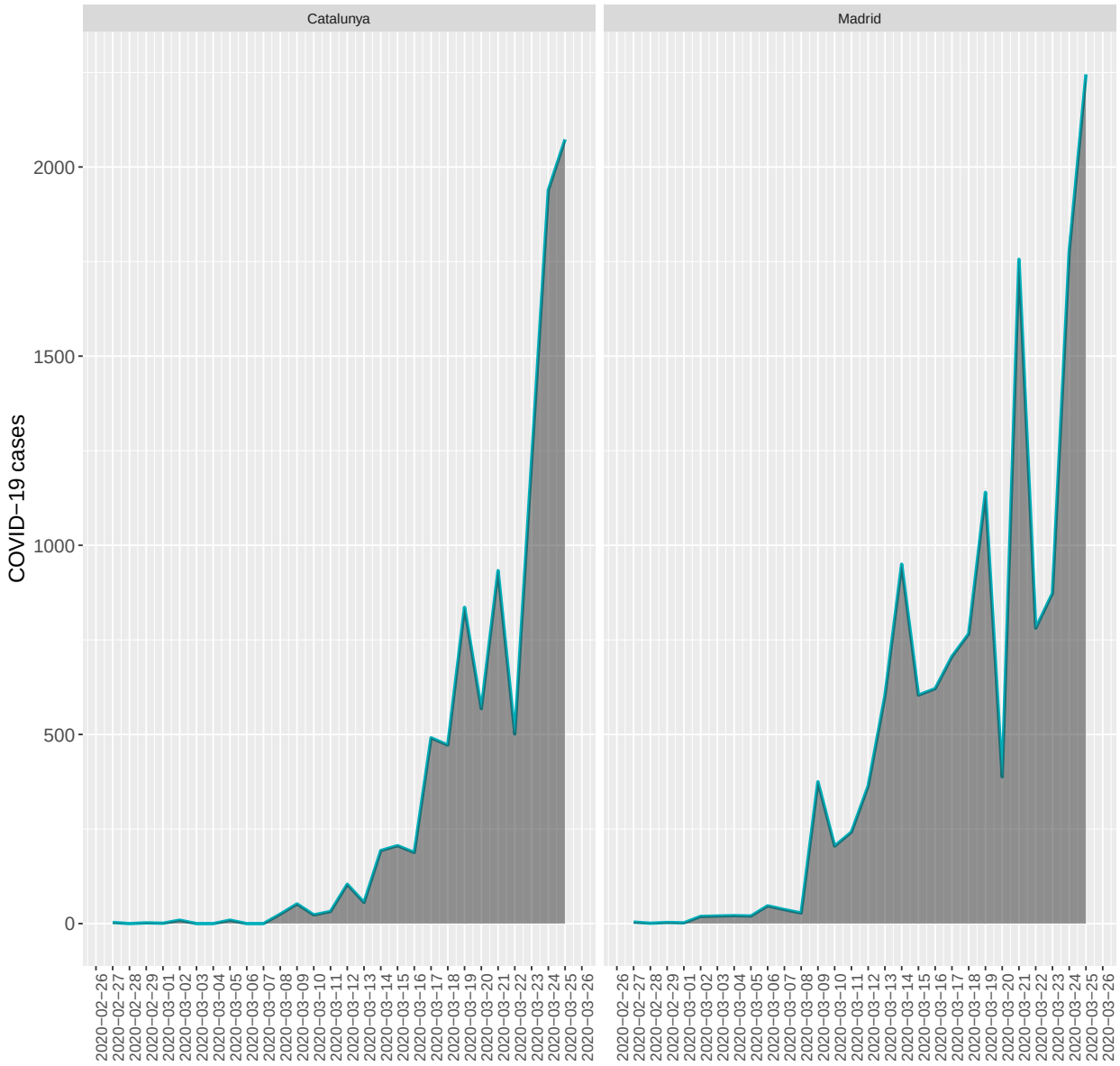

Figure 1 (c): Daily COVID−19 cases from 27−02−2020 to 25−03−2020

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

	minimum	mean	median	maximum	standard deviation	dispersion index
Andalucia	0.00	107.50	22.00	539.00	147.01	201.05
Aragon	0.00	30.64	13.00	149.00	42.54	59.04
Asturias	0.00	27.82	13.50	142.00	36.41	47.64
Baleares	0.00	20.07	2.50	85.00	29.18	42.43
Canarias	0.00	23.46	11.00	100.00	28.90	35.59
Cantabria	0.00	18.21	2.00	85.00	28.90	45.84
Castilla La Mancha	0.00	105.21	38.00	396.00	133.25	168.75
Castilla Leon	0.00	105.00	22.00	480.00	144.24	198.13
Catalunya	0.00	354.89	54.00	2073.00	569.67	914.44
Extremadura	0.00	26.50	7.00	143.00	38.10	54.77
Galicia	0.00	59.04	13.50	293.00	83.91	119.27
La Rioja	0.00	33.14	25.50	126.00	32.80	32.45
Madrid	1.00	521.32	369.50	2245.00	607.79	708.60
Murcia	0.00	17.04	7.00	92.00	22.94	30.88
Navarra	0.00	42.75	25.00	183.00	51.68	62.48
Pais Vasco	0.00	126.96	64.00	543.00	143.67	162.58
Valencia	0.00	103.39	14.00	449.00	131.63	167.59

Table 1: Summary of the daily COVID-19 cases from 27-02-20 to 25-03-2020 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		1.3612	0.18	0.6442	0.5461	336.7
s.e. (Andalucia)		0.1121	0.0045	0.0951	0.0254	
Aragon		0.1813	0.1738	0.3621	0.3311	226.8
s.e. (Aragon)		0.2022	0.0082	0.1127	0.0509	
Asturias		0.3133	0.1707	0.5809	0.4722	251.1
s.e. (Asturias)		0.2116	0.0086	0.1188	0.0347	
Baleares		-0.9076	0.1993	0.1521	0	193.5
s.e. (Baleares)		0.2386	0.0098	0.0803	NaN	
Canarias		0.2661	0.1558	0.2942	0.2626	174.5
s.e. (Canarias)		0.2429	0.01	0.1122	0.0774	
Cantabria		0.6085	0.1419	0.6643	0.1602	179.7
s.e. (Cantabria)		0.3241	0.0128	0.0924	0.0315	
Castilla La Mancha		2.1379	0.1405	0.4643	0.1156	458.6
s.e. (Castilla La Mancha)		0.1127	0.0047	0.0943	0.0203	
Castilla Leon		1.4623	0.1705	0.5447	0.2922	333.6
s.e. (Castilla Leon)		0.1813	0.0073	0.1016	0.0331	
Catalunya		1.4573	0.2231	0.5387	0.4255	565.1
s.e. (Catalunya)		0.0704	0.0028	0.0982	0.0161	
Extremadura		0.1661	0.1663	0.3982	0.0469	209.1
s.e. (Extremadura)		0.2373	0.0098	0.0985	0.0281	
Galicia		0.6713	0.1786	0.4514	0.0657	276.7
s.e. (Galicia)		0.1674	0.0069	0.0966	0.0243	
La Rioja		1.9767	0.1008	0.4858	0.3654	270.6
s.e. (La Rioja)		0.1377	0.006	0.0978	0.0316	
Madrid		3.7811	0.1446	0.5	0.3734	1278.3
s.e. (Madrid)		0.0364	0.0016	0.0945	0.0086	
Murcia		-1.0338	0.2002	0.2308	0.4793	150
s.e. (Murcia)		0.2652	0.011	0.1308	0.077	
Navarra		0.6814	0.1685	0.5971	0.6188	314
s.e. (Navarra)		0.1501	0.0061	0.1054	0.0321	
Pais Vasco		2.1253	0.1502	0.4768	0.537	449.8
s.e. (Pais Vasco)		0.0949	0.0039	0.0992	0.0248	
Valencia		4.0242	0.0644	0.6071	0.0729	504.6
s.e. (Valencia)		0.138	0.0058	0.0923	0.0086	

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

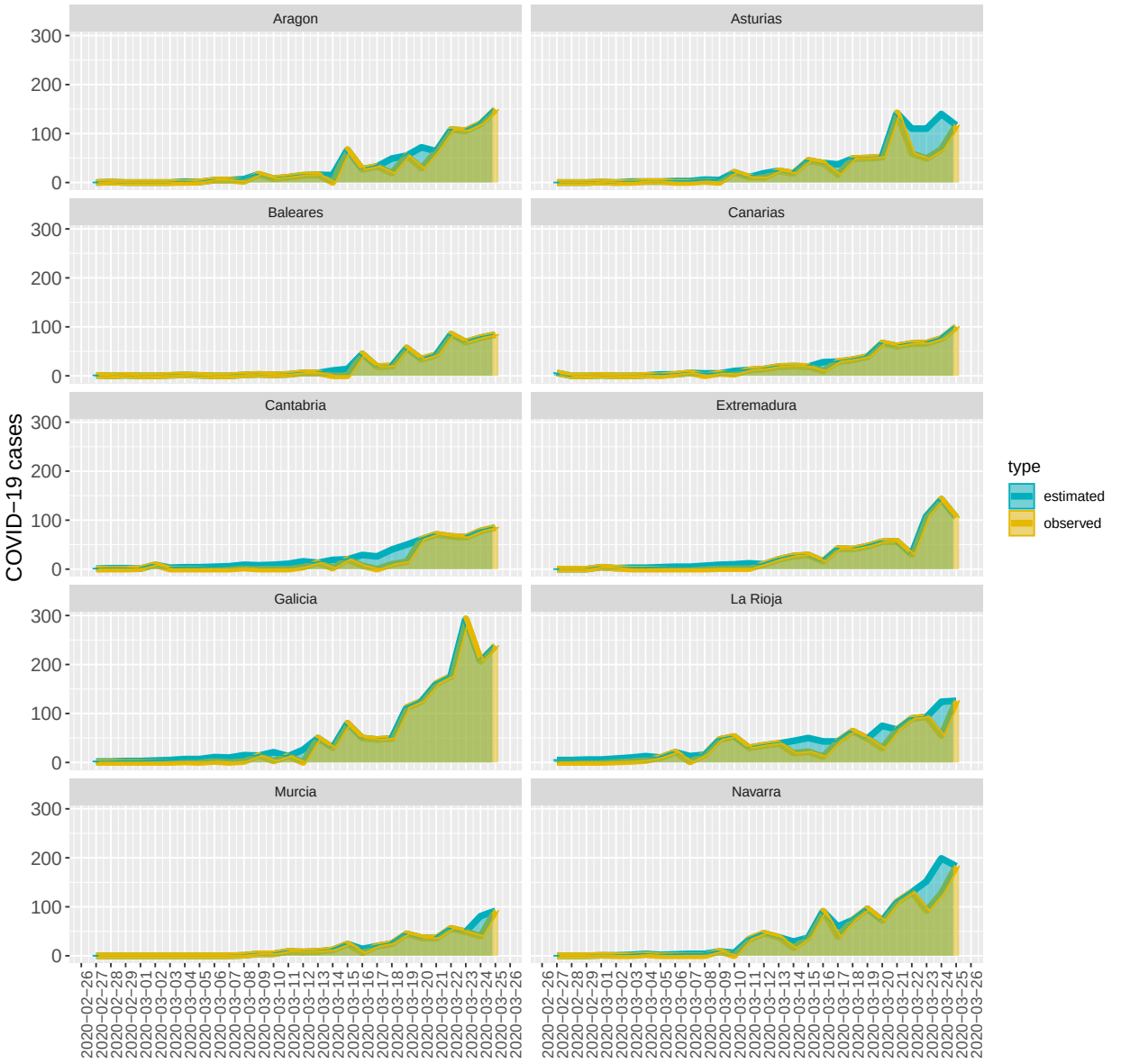

Figure 2 (a): Observed and truly daily cases

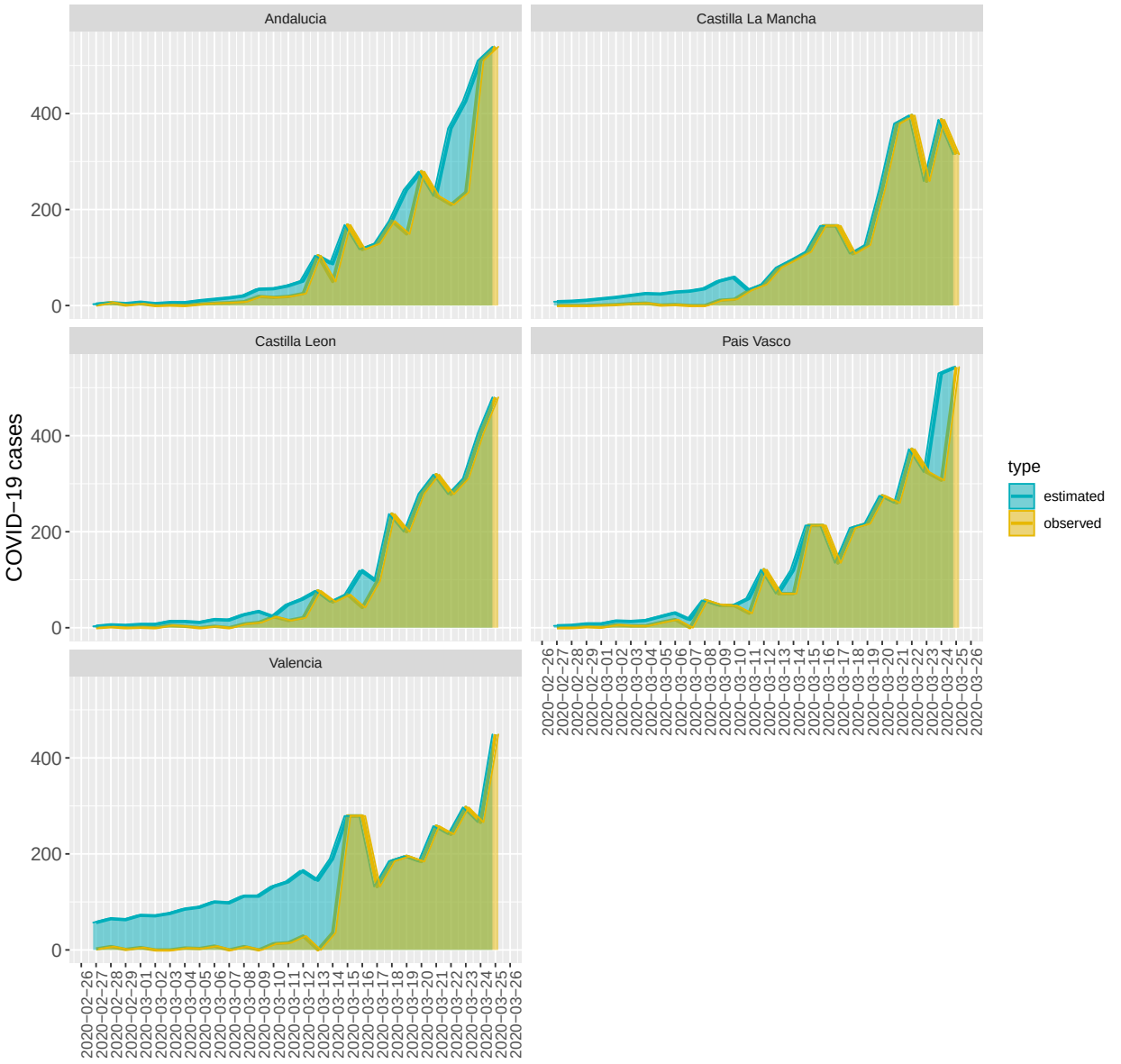

Figure 2 (b): Observed and truly daily cases

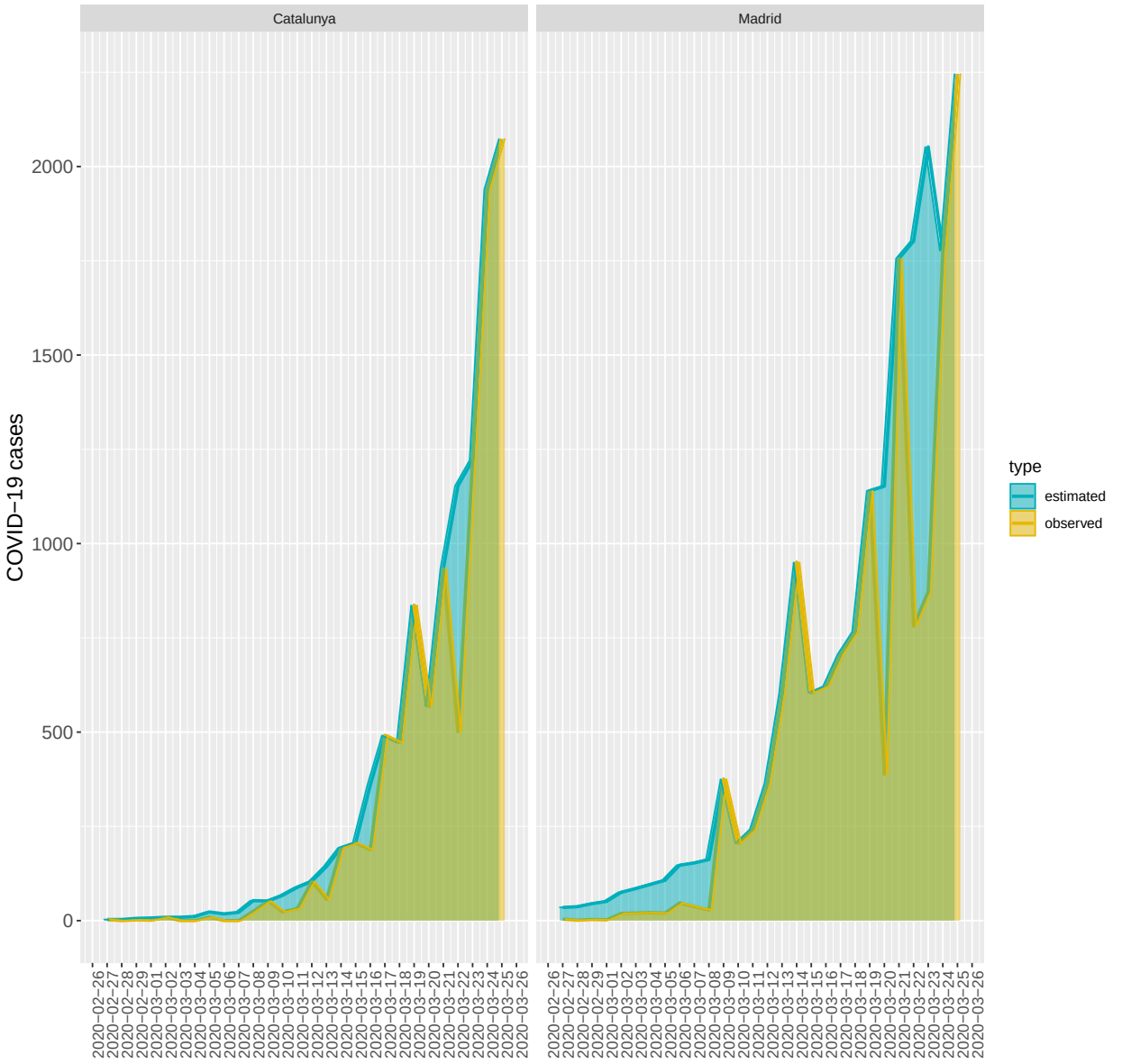

Figure 2 (c): Observed and estimated daily cases

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	107.50	129.61	17.06
Aragon	30.64	34.04	9.97
Asturias	27.82	36.07	22.87
Baleares	20.07	20.96	4.26
Canarias	23.46	24.64	4.78
Cantabria	18.21	25.50	28.57
Castilla La Mancha	105.21	115.68	9.05
Castilla Leon	105.00	114.82	8.55
Catalunya	354.89	395.07	10.17
Extremadura	26.50	28.46	6.90
Galicia	59.04	62.89	6.13
La Rioja	33.14	42.32	21.69
Madrid	521.32	655.43	20.46
Murcia	17.04	18.79	9.32
Navarra	42.75	49.50	13.64
Pais Vasco	126.96	141.11	10.02
Valencia	103.39	162.11	36.22

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

CCAA	estimated	observed	lethality-estimated	lethality-observed	mortality
Andalucia	3629.00	3010.00	3.11	3.75	1.34
Aragon	953.00	858.00	4.20	4.66	3.03
Asturias	1010.00	779.00	2.48	3.21	2.44
Baleares	587.00	562.00	2.21	2.31	1.13
Canarias	690.00	657.00	3.04	3.20	0.98
Cantabria	714.00	510.00	1.96	2.75	2.41
Castilla La Mancha	3239.00	2946.00	8.12	8.93	12.94
Castilla Leon	3215.00	2940.00	5.13	5.61	6.88
Catalunya	11062.00	9937.00	4.66	5.19	6.72
Extremadura	797.00	742.00	4.89	5.26	3.65
Galicia	1761.00	1653.00	1.53	1.63	1.00
La Rioja	1185.00	928.00	3.12	3.99	11.68
Madrid	18352.00	14597.00	9.94	12.50	27.39
Murcia	526.00	477.00	0.95	1.05	0.33
Navarra	1386.00	1197.00	2.38	2.76	5.04
Pais Vasco	3951.00	3555.00	3.92	4.36	7.02
Valencia	4539.00	2895.00	3.15	4.94	2.86

Table 4: Estimated mortality and lethality rates