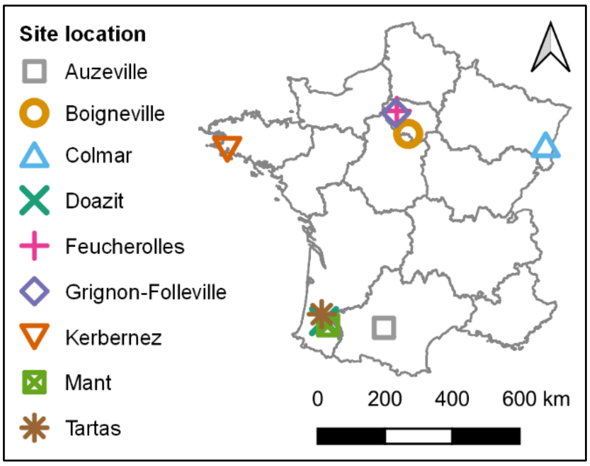
Supplementary Material

A robust initialization method for accurate soil organic carbon simulations

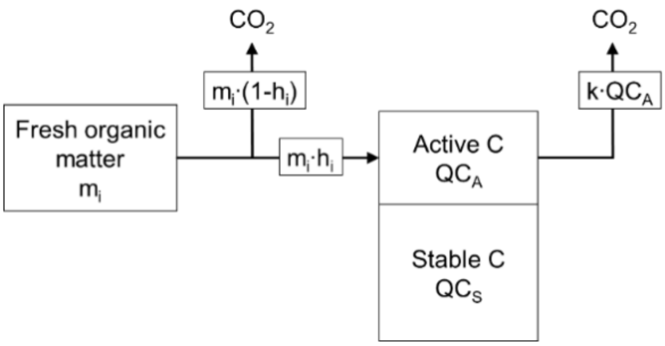
Eva Kanari[1,2], Lauric Cécillon[1,3], François Baudin[2], Hugues Clivot[4], Fabien Ferchaud[5], Sabine Houot[6], Florent Levavasseur[6], Bruno Mary[5], Laure Soucémarianadin[7], Claire Chenu[6], Pierre Barré[1]

[1]Laboratoire de Géologie, École normale supérieure, CNRS, Université PSL, IPSL, Paris, France
[2]ISTeP, UMR 7193 Sorbonne Université, CNRS, Paris, France
[3]Normandie Univ, UNIROUEN, INRAE, ECODIV, Rouen, France
[4]Université de Reims Champagne Ardenne, INRAE, FARE, UMR A 614, 51097 Reims, France
[5]BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université Picardie Jules Verne, F-02000, Barenton-Bugny, France
[6]UMR ECOSYS, INRAE, AgroParisTech, Université Paris-Saclay, Thiverval-Grignon, France
[7]ACTA - les instituts techniques agricoles, Paris, France

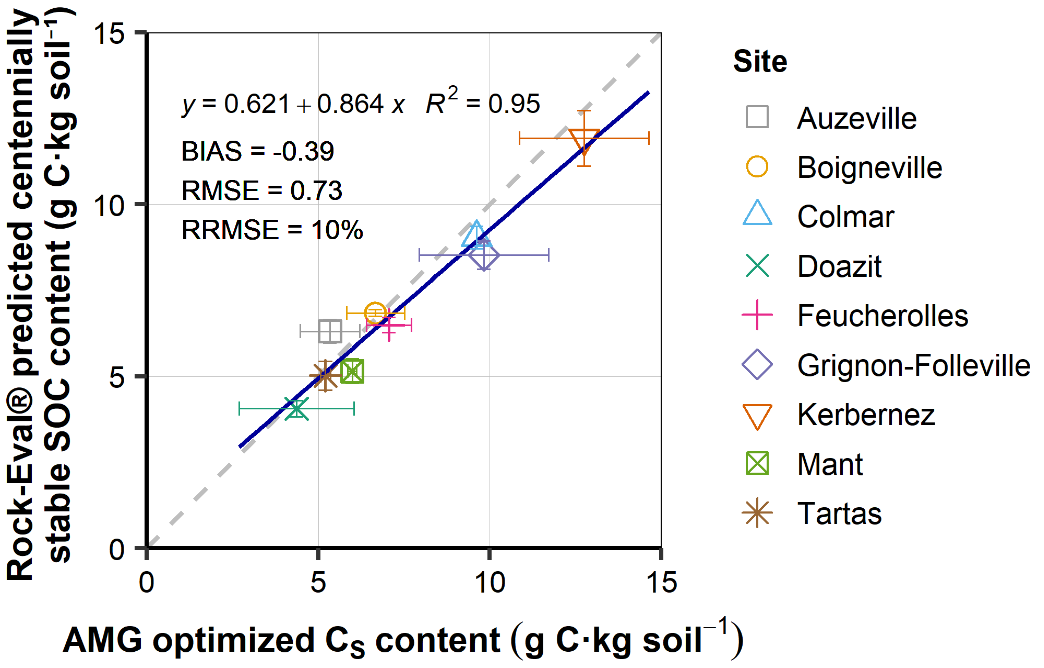
Correspondence to: Eva Kanari (kanari@geologie.ens.fr), Lauric Cécillon (lauric.cecillon@inrae.fr) and Pierre Barré (barre@biotite.ens.fr)



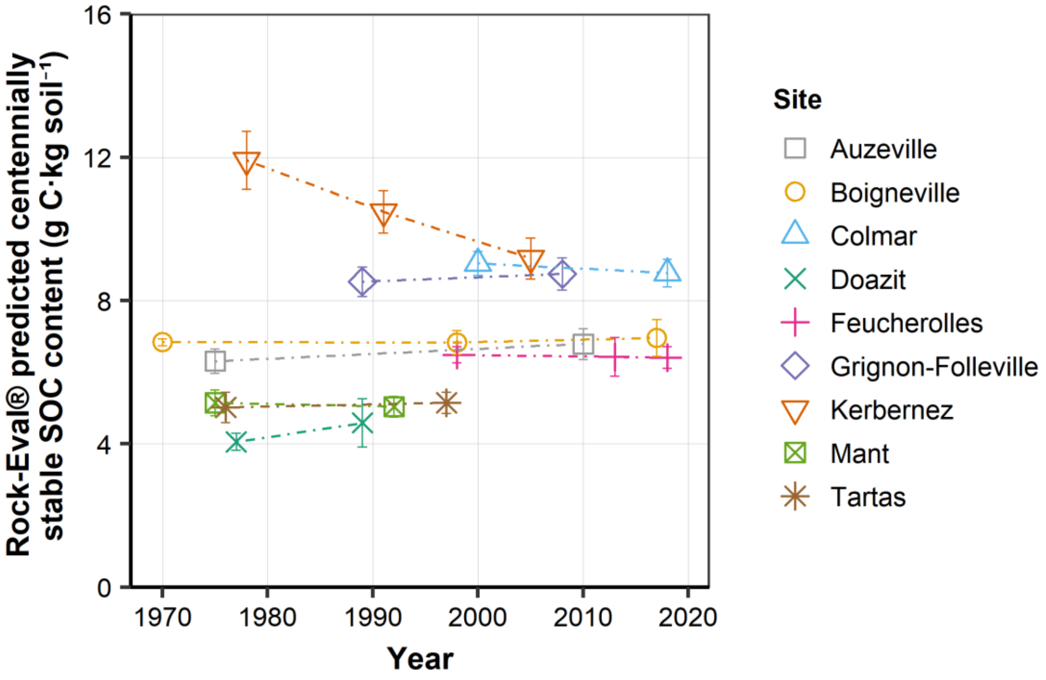
Supplementary Material Figure 1: Location of the nine French long-term agricultural experiments used in this study.



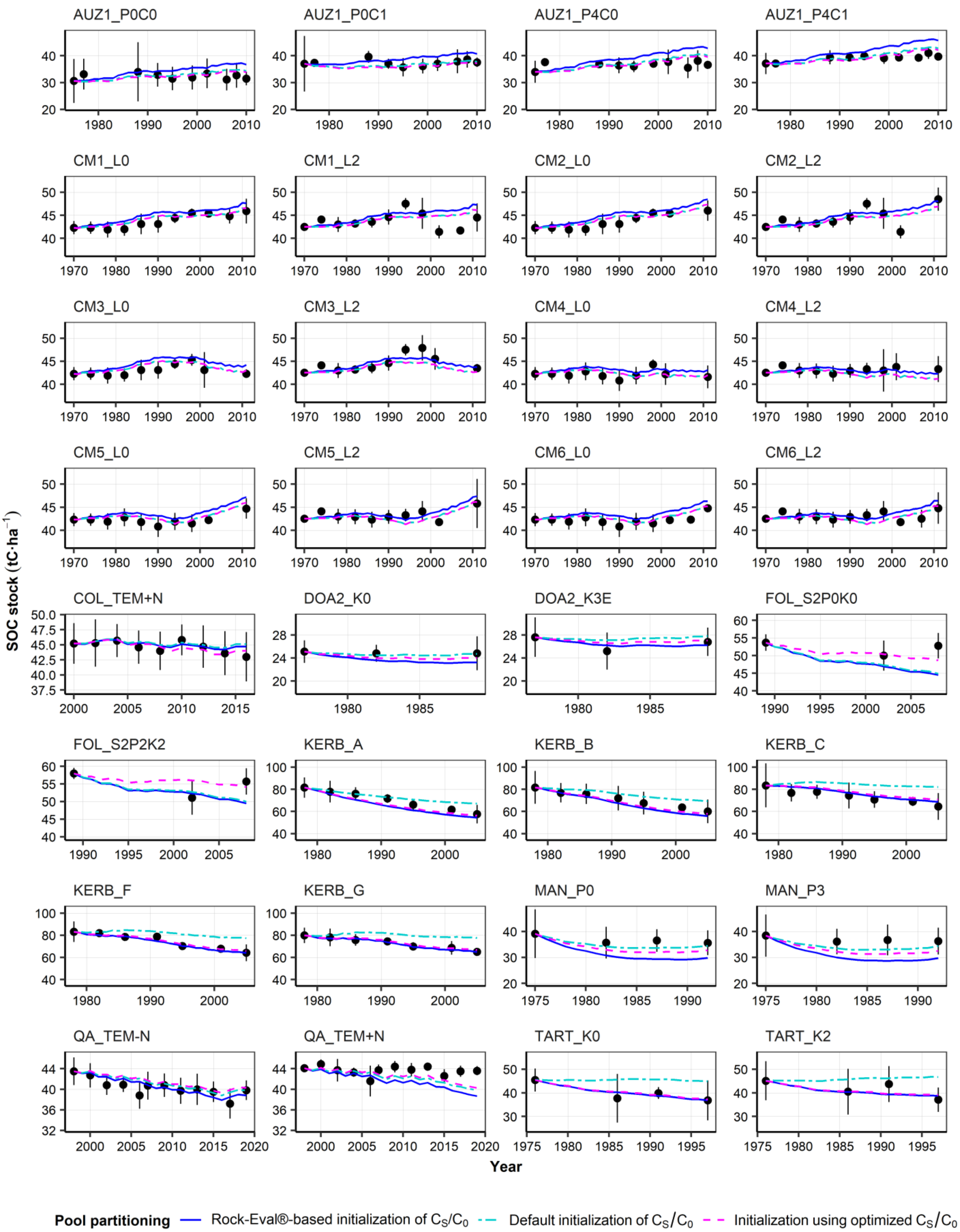
Supplementary Material Figure 2: Conceptual schematic diagram of the AMG model of SOC dynamics (modified from ref[43,80]), showing SOC pools, fluxes and transport rates. A fraction (1-h) of fresh organic matter (m) is yearly mineralized and released in the atmosphere, whereas a fraction (h) is incorporated into the active SOC pool (C_A). The coefficient of mineralization (k) controls carbon discharge from C_A into the atmosphere. There is no exchange with the stable SOC pool (C_S).



Supplementary Material Figure 3: Centennially stable SOC content predicted by the Rock-Eval®-based PARTY$_{SOC}$ machine-learning model compared to the AMG *ex-post* optimized stable SOC content. Points represent site-mean values based on initial topsoil samples. Statistics refer to the linear regression between x and y values (blue solid line). Horizontal error bars show the uncertainty associated with the optimal C$_S$ content, calculated as the standard deviation of treatment-wise optimized C$_S$ content. Vertical error bars represent the prediction error of the centennially stable SOC content values, calculated from the standard deviation of the PARTY$_{SOC}$ model predictions on initial topsoil samples.



Supplementary Material Figure 4: Centennially stable SOC content predicted by Rock-Eval® as a function of time of the experiment. The points on the plot represent mean values for the shown dates and the vertical error bars represent the standard deviation of the sample set used for averaging. The apparent decrease in centennially stable SOC content for the site of Kerbernez could be explained by changes in soil bulk density, caused by the change in land-use (form grassland to cropland) in 1958. The subsequent soil compaction may have led to inclusion of deeper soil during standard sampling of the 0–25cm layer, causing a false effect of SOC content decrease. Lack of regular soil bulk density measurements during the experiment (1978–2005) hinders explicit analysis of this hypothesis.



Supplementary Material Figure 5: AMG simulations of observed SOC dynamics for the 32 treatments used in this study. The black points represent observed SOC stocks in topsoils, the vertical error bars indicate the confidence interval of the measurements, and each line corresponds to a simulation resulting from a different initialization method, namely initial pool partitioning according to: Rock-Eval®-based SOC pool partitioning in blue, AMG default C_S/C_0 in cyan, and AMG ex-post optimized C_S/C_0 in magenta. Note the different y-axis range across sites. The treatment IDs and their corresponding sites are presented in Supplementary Material Table 1.

Supplementary Material Table 1: Information on site location, long-term land cover history, climate and soil characteristics. Note that the arable land cover class may include temporary grassland in crop rotations, while the grassland land cover class does not include cultivated crops.

	Auzeville	**Boigneville**	**Colmar**	**Doazit**	**Feucherolles**	**Grignon-Folleville**	**Kerbernez**	**Mant**	**Tartas**
Latitude ° N	43.527479	48.327843	48.059271	43.700824	48.896501	48.841722	47.946698	43.5917	43.865475
Longitude ° E	1.506059	2.382406	7.328160	−0.629406	1.972125	1.936675	−4.127084	−0.5028	−0.729405
∗Historical land cover 1820–1866	arable land	arable land	arable land	arable land	arable land	arable land	arable land	arable land	grassland
†Historical land cover 1950–1965	arable land	arable land	arable land	arable land	arable land	arable land	grassland	arable land	arable land
‡Treatment	AUZ1_P0C0 AUZ1_P0C1 **AUZ1_P4C0** **AUZ1_P4C1**	**CM1_L0** **CM1_L2** CM2_L0 CM2_L2 CM3_L0 CM3_L2 CM4_L0 CM4_L2 CM5_L0 CM5_L2 CM6_L0 CM6_L2	**TEM+N**	**DOA2_K0** DOA2_K3E	**QU_TEM-N** **QU_TEM+N**	**FOL_S2P0K0** **FOL_S2P2K2**	**KERB_A** **KERB_B** **KERB_C** KERB_F KERB_G	**MAN_P0** MAN_P3	**TART_K0** TART_K2
§MAT (°C)	13.5	10.9	11.2	13.1	10.8	11.0	11.9	13.1	13.4
‖MAP-PET (mm)	−290	−87	−222	384	5	−69	489	364	383
Bulk density (g · cm^{-3})	1.40	1.44	1.30	1.40	1.38	1.40	1.30	1.40	1.40
Clay (g · kg soil^{-1})	275	248	180	72	170	244	163	94	43
Silt (g · kg soil^{-1})	339	672	628	403	779	601	391	554	166
Sand (g · kg soil^{-1})	372	80	76	525	51	97	446	349	791
$CaCO_3$ (g · kg soil^{-1})	15	0	115	0	0	58	0	3	0
C:N ratio	8.0	9.0	9.2	10.6	9.3	9.8	11.4	9.4	13.0
pH	7.6	6.8	8.3	6.4	6.9	8.1	5.7	7.6	6.0
Reference	ref[46]	ref[47]	ref[48]	ref[49]	ref[50]	ref[51]	ref[52]	ref[53]	ref[54]

∗ French "Carte de l'Etat Major", IGN
† aerial photography, IGN
‡ Treatments from which samples were available for Rock-Eval® analysis are in bold
§ Mean annual temperature
‖ Mean annual precipitation-potential evapotranspiration

Supplementary Material Table 2: Measurement error and variation of initial SOC stock values, and variation of initial centennially stable SOC proportion amongst sites. Left part: Comparison of the variation (standard deviation) and uncertainty (confidence interval) associated with initial SOC stock measurements. Right part: variation of initial centennially stable SOC proportions predicted by the $PARTY_{SOC}$ machine-learning model for each site.

Site	Initial SOC stock ($tC \cdot ha^{-1}$)			Initial centennially stable SOC proportion predicted using the $PARTY_{SOC}v2.0_{EU}$ statistical model	
	Mean	**Standard deviation**	**Confidence interval**	**Mean**	**Standard deviation**
Auzeville	34.68	2.66	13.30	0.74	0.01
Boigneville	42.40	0.10	2.30	0.68	0.05
Colmar	45.20	-	6.74	0.64	0.02
Doazit	26.35	1.25	5.38	0.57	0.01
Grignon-Folleville	55.85	2.15	3.93	0.64	0.04
Feucherolles	43.80	0.42	3.49	0.62	0.02
Kerbernez	81.98	1.29	24.01	0.44	0.02
Mant	38.75	0.35	17.55	0.52	0.05
Tartas	45.25	0.15	13.14	0.44	0.05