



# GlobEmission

# ESA Contract No 4000104001/11/I-NB

Algorithm Theoretical Baseline Document Version 3.4

Date of issue: 31 March 2016



	Name
Lead Author	Ronald van der A
Contributing Authors	Jenny Stavrakou, Maite Bauwens, Nele Veldeman, Bino Maiheu, Lyana Curier, Mikhail Sofiev, Gerrit de Leeuw, Edith Rodriguez; Julius Vira, Bas Mijling
Reviewed by	Claus Zehner
Distribution list	Claus Zehner, ESA Public

Document Change Record					
Date	Issue		Pages affected	Description	
9 January 2012	draft		All	Template	
4 April 2012	1.0		All	First issue	
30 October 2012	2.0		52	Small updates	
20 December 2013	2.2		52-54	Description new algorithm for spatial disaggregation	
16 September 2014	3.0		All	New version for CCN1	
10 December 2014	3.1			Updates	
17 March 2015	3.2		42-49	Update of FMI algorithm description	
31 March 2016	3.4		Section 4.2 and 4.7	Update of algorithm descriptions of VITO and KNMI (DECSO)	



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:3 of 64



# **Table of Contents**

1	Introdu	uction	6
	1.1 Pu	rpose and scope	6
2	Docum	nents	6
	2.1 Re	ference documents	6
	2.2 In:	formative documents	12
3	Overv	iew of algorithms	13
4	Algori	thm descriptions	14
	4.1 G	lobal Emission Estimates	14
	4.1.1	Introduction	14
	4.1.2	Input data	14
	4.1.3	HCHO simulation using the IMAGESv2 global CTM	15
	4.1.4	Error analysis	18
	4.1.5	References	19
	4.2 Re	gional Emission Estimates	21
	4.2.1	Input data	21
	4.2.2	Algorithm description	21
	4.2.3	Error analysis	
	4.3 Er	nission Inventories for European products	
	4.3.1	Input data	
	4.3.2	Algorithm description	
	4.3.3	Error analysis	43
	4.4 Er	nission Estimates related to Aerosols	44
	4.4.1	Input data	44
	4.4.2	Algorithm description	44
	4.4.3	Error analysis	46
	4.5 Er	nission Estimates related to fires	
	4.5.1	Input data	
	4.5.2	Algorithm description	
	4.5.3	Error analysis	50
	4.6 Er	nission Estimates related to fires	52
	4.6.1	Input data	



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:5 of 64

6.2 Algorithm description	4.6.2
6.3 Error analysis	4.6.3
High Resolution Emission Inventories	4.7 Hig
6.1 Input data56	4.6.1
6.2 Algorithm description	4.6.2
7.1 Error analysis	4.7.1



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:6 of 64

# **1** Introduction

# 1.1 Purpose and scope

The algorithms used in the services of GlobEmission to derive emission estimates from satellite data are described in detail in this document. Each of the 5 algorithms described has been developed for a specific type of emission and spatial scale. The performance of the algorithms in combination with the available satellite data has been tested. This document contains for each algorithm a description of the input data, detailed algorithm description and an error analysis. The error analysis is based on theoretical calculations but tested with real satellite data.

# 2 Documents

# 2.1 Reference documents

- [RD-3] Andreae, M. O. and Merlet, P.: Emission of trace gases and aerosols from biomass burning, Global Biogeochem. Cy., 15, 955–966, 2001, and updates from Andreae, M. O., personal communication, 2007.
- [RD-4] Arey, J., Obermeyer, G., Aschmann, S. M., Chattopadhyay, S., Cusick, R. D., and Atkinson, R.: Dicarbonyl Products of the OH Radical-Initiated Reaction of a Series of Aromatic Hydrocarbons, Environ. Sci. Technol., 43(3), 683–689, doi:10.1021/es8019098, 2009.
- [RD-5] Butler, T. M., Taraborrelli, D., Brühl, C., Fischer, H., Harder, H., Martinez, M., Williams, J., Lawrence, M. G., and Lelieveld, J., Improved simulation of isoprene oxidation chemistry with the ECHAM5/MESSy chemistry-climate model: lessons from the GABRIEL airborne field campaign, Atmos. Chem. Phys., 8, 4529–4546, 2008.
- [RD-6] Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G. R.: The Kinetic PreProcessor KPP A Software Environment for Solving Chemical Kinetics, Comput. Chem. Eng., 26,1567–1579, 2002.
- [RD-7] Feierabend, K. J., Zhu, L., Talukdar, R. K., and Burkholder, J. B.: Rate Coefficients for the OH+HC(O)C(O)H (Glyoxal) Reaction between 210 and 390 K, J. Phys. Chem. A, 112(1), 73–82, 2008.
- [RD-8] Fu, T.-M., Jacob, D. J., Wittrock, F., Burrows, J. P., Vrekoussis, M., and Henze, D. K.: GlobalFF budgets of atmospheric glyoxal and methylglyoxal, and implications for formation of secondary organic aerosols, J. Geophys. Res., 113, D15303, doi:10.1029/2007JD009505, 2008.
- [RD-9] Giering, R.: Tangent linear and adjoint biogeochemical models, in: Inverse methods in Global Biogeochemical Cycles, Geophysical Monograph, 114, 2000.
- [RD-10] Giering, R. and Kaminski, T.: Recipes for adjoint code construction, ACM Trans. On Math. Software, 24, 437–474, 1998.
- [RD-11] Gilbert, J.-C. and Lemaréchal, C.: Some numerical experiments with variable storage quasi-Newton algorithms, Math. Programming, 45, 407–435, 1989.
- [RD-12] Lelieveld, J., Butler, T. M., Crowley, J. N., Dillon, T. J., Fischer, H., Ganzeveld, L., Harder, H., Lawrence, M. G., Martinez, M., Taraborrelli, D., and Williams, J.: Atmospheric oxidation capacity sustained by a tropical forest, Nature, 452, 737–740, 2008.



- [RD-13] Li, Q., Jacob, D. J., Yantosca, R. M., Heald, C. L., Singh, H. B., Koike, M., Zhao, Y., Sachse, G. W., and Streets, D. G.: A global three-dimensional model analysis of the atmospheric budgets of HCN and CH3CN: constraints from aircraft and ground measurements, J. Geophys. Res., 108(D21), 8827, doi:10.1029/2002JD003075, 2003.
- [RD-14] Müller, J.-F. and Stavrakou, T.: Inversion of CO and NOx emissions using the adjoint of the IMAGES model, Atmos. Chem. Phys., 5, 1157–1186, 2005.
- [RD-15] Müller, J.-F., Stavrakou, T., Wallens, S., De Smedt, I., Van Roozendael, M., Potosnak, M. J., Rinne, J., Munger, B., Goldstein, A., and Guenther, A. B.: Global isoprene emissions estimated using MEGAN, ECMWF analyses and a detailed canopy environment model, Atmos. Chem. Phys., 8, 1329–1341, 2008.
- [RD-16] Ohara, T., H. Akimoto, J. Kurokawa, N. Horii, K. Yamaji, X. Yan, and T. Hayasaka, An Asian emission inventory of anthropogenic emission sources for the period 1980–2020, Atmos. Chem. Phys., 7, 4419-4444, 2007.
- [RD-17] Olivier, J. G. J., Berdowski, J. J. M., Peters, J. A. H. W., Bakker, J., Visschedijk, A. J. H., and Bloos, J.-P. J.: Applications of EDGAR, Including a description of EDGAR 3.0: reference database with trend data for 1970–1995, RIVM report no. 773301 001/ NOP report no. 410200 051, RIVM, Bilthoven, 2001.
- [RD-18] Olivier, J. G. J.: Part III: Greenhouse gas emissions. 1. Shares and trends in greenhouse gas emissions;
   2. Sources and methods: greenhouse gas emissions for 1990 and 1995 in "CO<sub>2</sub> emissions from fuel combustion 1971–2000", International Energy Agency, Paris, ISBN 92-64-09794-5, 1–31, 2002.
- [RD-19] Palmer, P., Abbot, D. S., Fu, T.-M., Jacob, D. J., Chance, K., Kurosu, T. P., Guenther, A., Wiedinmyer, C., Stanton, J. S., Pilling, M. J., Pressley, S. N., Lamb, B., and Summer, A. L.: Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, J. Geophys. Res., 111, D12315, doi:10.1029/2005JD006689, 2006.
- [RD-20] Peeters, J., Vereecken, L., and Fantechi, G.: The detailed mechanism of the OH-initiated atmospheric oxidation of alpha-pinene: a theoretical study, Phys. Chem. Chem. Phys., 3, 5489–5504, 2001.
- [RD-21] Peeters, J. and Müller, J.-F.: HOx radical regeneration in isoprene oxidation via peroxy radical isomerisations, II: Experimental evidence and global impact, Phys. Chem. Chem. Phys., doi:10.1039/C0CP00811G, 2010.
- [RD-22] Sander, S. P., Finlayson-Pitts, B. J., Friedl, R. R., Golden, D. M., Huie, R. E., Keller-Rudek, H., Kolb, C. E., Kurylo, M. J., Molina, M. J., Moortgat, G. K., Orkin, L. V., Ravishankara, A. R., and Wine, P. H.: Chemical Kinetics and Photochemical data for use in atmospheric studies, Evaluation number 15, NASA Panel for data evaluation, JPL Publication 06-2, Jet Propulsion Laboratory, Pasadena, 2006.
- [RD-23] Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, Atmos. Chem. Phys., 6, 187–195, 2006, http://www.atmos-chemphys.net/6/187/2006/.
- [RD-24] Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of nonaromatic volatile organic compounds, Atmos. Chem. Phys., 3, 161–180, 2003.
- [RD-25] Stavrakou, T., J. Peeters, and J.-F. Müller : Improved global modelling of HOx recycling in isoprene oxidation: evaluation against the GABRIEL and INTEX-A aircraft campaign measurements, Atmos. Chem. Phys., 10, 9863-9878, 2010.
- [RD-26] Stavrakou, T., A. Guenther, A. Razavi, L. Clarisse, C. Clerbaux, P.-F. Coheur, D. Hurtmans, F. Karagulian, M. De Mazière, C. Vigouroux, C. Amelynck, N. Schoon, Q. Laffineur, B. Heinesch, M. Aubinet, C. Rinsland, and J.-F. Müller: First space-based derivation of the global atmospheric methanol emission fluxes, Atmos. Chem. Phys., 11, 4873–4898, 2011.
- [RD-27] Taraborrelli, D., M. G. Lawrence, T. M. Butler, R. Sander, and J. Lelieveld : Mainz Isoprene Mechanism 2 (MIM2): an isoprene oxidation mechanism for regional and global atmospheric modeling, Atmos. Chem. Phys., 9, 2751-2777, 2009.



- [RD-28] Tyndall, G. S., Orlando, J. J., Wallington, T. J., and Hurley, M. D.: Products of the chlorine-atom- and hydroxyl-radical-initiated oxidation of CH<sub>3</sub>CN, J. Phys. Chem. A, 105, 5380–5384, 2001.
- [RD-29] van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, M. Mu, P. S. Kasibhatla, D. C. Morton, R. S. DeFries, Y. Jin, and T. T. van Leeuwen: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, 2010.
- [RD-30] Volkamer, R., Platt, U., and Wirtz, K.: Primary and Secondary Glyoxal Formation from Aromatics: Experimental Evidence for the Bicycloalkyl-Radical Pathway from Benzene, Toluene, and p-Xylene, J. Phys. Chem. A, 105, 7865–7874, 2001.
- [RD-31] Volkamer, R., Spietz, P., Burrows, J. P., and Platt, U.: High resolution absorption cross section of Glyoxal in the UV/vis and IR spectral ranges, J. Photoch. Photobio. A, 172, 35–46, doi:10.1016/j.jphotochem.2004.11.011, 2005a.
- [RD-32] Volkamer, R., Barnes, I., Platt, U., Molina, L. T., and Molina, M. J.: Remote Sensing of Glyoxal by Differential Optical Absorption Spectroscopy (DOAS): Advancements in Simulation Chamber and Field Experiments, NATO Sciences Series, IV, Earth and Environmental Sciences, "Environmental Simulation Chambers: Application to Atmospheric Chemical Processes", edited by: Barnes, I. and Rudzinski, K. J., Kluwer Academic Publishers, Dordrecht, The Netherlands, ISBN: 1-4020-4231-0, Vol. 62, 2005b.

- [RD-1] Bessagnet, B., A. Hodzic, R. Vautard, M. Beekmann, S. Cheinet, C. Honoré, C. Liousse, and L. Rouïl (2004), Aerosol modeling with CHIMERE—Preliminary evaluation at the continental scale, Atmos. Environ., 38, 2803–2817.
- [RD-2] Blond, N., K. F. Boersma, H. J. Eskes, R. J. van der A, M. Van Roozendael, I. De Smedt, G. Bergametti, and R. Vautard (2007), Intercomparison of SCIAMACHY nitrogen dioxide observations, in situ measurements and air quality modeling results over Western Europe, J. Geophys. Res., 112, D10311, doi:10.1029/2006JD007277.
- [RD-3] Boersma, K. F., D. J. Jacob, H. J. Eskes, R. W. Pinder, J. Wang, and R. J. van der A (2008), Intercomparison of SCIAMACHY and OMI tropospheric NO<sub>2</sub> columns: observing the diurnal evolution of chemistry and emissions from space, J. Geophys. Res., 113, D16S26, doi:10.1029/2007JD008816
- [RD-4] Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D. (2011): An improved tropospheric NO<sub>2</sub> column retrieval algorithm for the Ozone Monitoring Instrument, *Atmos. Meas. Tech.*, 4, 1905-1928, doi:10.5194/amt-4-1905-2011.
- [RD-5] Ding, J., R.J. van der A, B. Mijling, P.F. Levelt, and N. Hao (2015). NOx emission estimates during the 2014 Youth Olympic Games in Nanjing. Atmos. Chem. Phys. Discuss., 15, 6337-6372, doi:10.5194/acpd-15-6337-2015.[RD-6] Huijnen, V., J. Williams, M. van Weele, T. van Noije, M. Krol, F. Dentener, J. de Laat, F. Boersma, P. van Velthoven, P. Le Sager, H. Eskes, F. Alkemade, R. Scheele, S. Houweling and co-authors (2010) The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0, *Geoscientific Model Development*, 3, 445-473, doi:10.5194/gmd-3-445-2010.
- [RD7] Lehoucq, R. B.; Sorensen, D. C.; Yang, C. (1998), ARPACK Users Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. Philadelphia: SIAM. ISBN 978-0898714074.
- [RD-8] Martin, R. V., D. J. Jacob, K. Chance, T. P. Kurosu, P. I. Palmer, and M. J. Evans (2003), Global inventory of nitrogen oxide emissions constrained by space-based observations of NO<sub>2</sub> columns, J. *Geophys. Res.*, 108(D17), 4537, doi:10.1029/2003JD003453.



- [RD-9] Mijling, B., R.J. van der A, K.F. Boersma, M. Van Roozendael, I. De Smedt and H.M. Kelder, Reduction of NO<sub>2</sub> detected from space during the 2008 Beijing Olympic Games, *Geophys. Res. Lett.*, 2009, 36, doi:10.1029/2009GL038943.
- [RD-10] Mijling, B., R.J. van der A, R. J., and Q. Zhang (2013). Regional nitrogen oxides emission trends in East Asia observed from space. Atmos. Chem. Phys., 13, 12003-12012, doi:10.5194/acp-13-12003-2013.
- [RD-11] Mijling, B. and R. van der A (2012). Using daily satellite observations to estimate emissions of shortlived air pollutants on a mesoscopic scale. J. Geophys. Res. 117, D17, doi:10.1029/2012JD017817.
- [RD-12] Miyazaki, K., H.J. Eskes, and K. Sudo (2012), Global NO<sub>x</sub> emission estimates derived from an assimilation of OMI tropospheric NO<sub>2</sub> columns, *Atmos. Chem. Phys.*, 12, 2263-2288, doi:10.5194/acp-12-2263-2012, 2012.
- [RD-13] Schmidt, H., C. Derognat, R. Vautard, and M. Beekmann (2001), A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe, Atmos. Environ., 36, 6277– 6297.
- [RD-14] Vautard, R., P. H. J. Builtjes, P. Thunis, C. Cuvelier, M. Bedogni, B. Bessagnet, C. Honoré, N. Moussiopoulos, G. Pirovano, M. Schaap, R. Stern, L. Tarrason and P. Wind (2006), Evaluation and intercomparison of ozone and PM10 simulations by several chemistry transport models over 4 European cities within the CityDelta project, Atmos. Environ., 41, 173–188.
- [RD-15] Zhang, L, D. J. Jacob, K. F. Boersma, D. A. Jaffe, J. R. Olson, K. W. Bowman, J. R. Worden, A. M. Thompson, M. A. Avery, R. C. Cohen, J. E. Dibb, F. M. Flock, H. E. Fuelberg, L. G. Huey, W. W. McMillan, H. B. Singh, and A. J. Weinheimer (2008), Transpacific transport of ozone pollution and the effect of recent Asian emission increases on air quality in North America: an integrated analysis using satellite, aircraft, ozonesonde, and surface observations, *Atmos. Chem. Phys.*, 8, 6117-6136.
- [RD-16] Zhang, Q., D. G. Streets, G. R. Carmichael, K. He, H. Huo, A. Kannari, Z. Klimont, I. Park, S. Reddy, J. S. Fu, D. Chen, L. Duan, Y. Lei, L. Wang, and Z. Yao (2009), Asian emissions in 2006 for the NASA INTEX-B mission, *Atmos. Chem. Phys. Discuss.*, 9, 4081-4139.

- [RD-1] Baldocchi, D., Hicks B. B., and P. Camara. A canopy stomatal resistance model for gaseous deposition to vegetated surfaces. Atmos. Environ., 21: 91–101, 1988.
- [RD-2] Emberson, L., Ashmore M.R., Cambridge H.M., Simpson D., and J.P. Touvinen. Modellingstomatal ozone flux across europe. Environ. Poll., 109, No. 3:403–414, 2000.
- [RD-3] Emmons, L.K. ,Walters S., and Hess P. G. et al. Description and evaluation of the model for ozone and related chemical tracers, version 4 (mozart-4). Geosci. Model Dev., 3:43–67, 2010.
- [RD-4] Evensen, G., Sequential data assimilation with a nonlinear quasigeostrophic model using montecarlo methods to forecast error statistics. J. Geophys. Res., 99:10143–10162, 1994.
- [RD-5] Felzer, B., Cronin T., Reilly J.M., Melillo J.M., and Wang X. Impacts of ozone on trees and crops. Comptes Rendus Geosciences, 339, 2007.
- [RD-6] Inness A., Flemming, J., Flentje H., Huijnen V., Moinat P., Schultz M. G., and O.: Stein. Coupling global chemistry transport models to ecmwf's integrated forecast system. Geosci. Model Dev., 2:253– 265, 2009.
- [RD-7] Houtekamer, P.L. and H.L. Mitchell. A sequential ensemble kalman filter for atmospheric data assimilation. Mon. Weather. Rev., 129:123–137, 2001.
- [RD-8] Kaiser, J.W., M. Suttie, J. Flemming, J.-J. Morcrette, O. Boucher, and M.G. Schultz. Global real-time fire emission estimates based on space-borne fireradiative power observations. In AIP Conf. Proc. Vol. 1100, 2009.



- [RD-9] Kalman, R. E. A new aproach to linear filter and prediction theory. Journal of Basic Engineering., 82D:35–45, 1960.
- [RD-10] Kuenen, J., Denier van der Gon, H. Visschedijk A., van de Brugh H., and Gijlswijk R. Macceuropean emission inventory for the years 2003-2007. Technical report, report TNO-060-UT-2011-00, TNO, Utrecht, the Netherlands, 2011.
- [RD-11] Monahan, E. C., D. E. Spiel, and K. L. Davidson. A model of marine aerosol generation via whitecaps and wave disruption. Oceanic Whitecaps, pages 167–174, 1986. URL www.scopus.com. Cited By (since 1996): 203.
- [RD-12] Nenes, A., Pandis S.N., and Pilinis C. Isorropia: A new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols. Aquat.Geoch., 4:123–152, 1998.
- [RD-13] Schaap, M., and H.A.C. Denier van der Gon. On the variability of black smoke and carbonaceous aerosols in the netherlands. Atmospheric Environment, 41(28):5908–5920, 2007. URL www.scopus.com. Cited By (since 1996): 7.
- [RD-14] Schaap, M., M. van Loon, H. M. ten Brink, F. J. Dentener, and P. J. H. Builtjes. Secondary inorganic aerosol simulations for europe with special attention to nitrate. Atmospheric Chemistry and Physics, 4(3):857–874, 2004
- [RD-15] Schaap, M., Roemer M., Sauter M., Boersen G., Timmermans R., and Builtjes P.J.H. Lotos-euros documentation. Technical 609 report, TNO Report B&O 2005/297, 2005.
- [RD-16] Schaap, M., R. M. A. Timmermans, M. Roemer, G. A. C. Boersen, P. J. H. Builtjes, F. J. Sauter, G. J. M. Velders, and J. P. Beck. The lotoseuros model: Description, validation and latest developments. International Journal of Environment and Pollution, 32(2):270–290, 2008.
- [RD-17] Segers, A.J., H.J. Eskes, R.J. van der A, R.F. van Oss, and P.F.J. van Velthoven. Assimilation of gome ozone profiles and a global chemistrytransport model using a kalman filter with anisotropic covariance. Q.J.R.meteorologyrol. Soc., 131(606):477–502, januari 2005. URL http://www.rmets.org/publication/OJ/qirms05.php.
- [RD-18] Simpson, D., Fagerli H., Jonson J.E., Tsyro S., Wind P., and Tuovinen JP. Transboundary acidification, eutrophication and ground level ozone ineurope, part 1: unified emep model description. Technical report, EMEP Report 1/2003, Norwegian meteorologyrologicalInsitute, 2003.
- [RD-19] Swart, D.P.J., S. Jongen, H. Eskes, M. Schaap, R.M.A. Timmermans, A. Segers, A.M.M. Manders, F.J. Sauter, J.P.J. Berkhout, and D.E. Lolkema. Smogprog: Towards operational smog forecasts using nearrealtime satellite measurements. Technical report, Final report NIVR project SMOGPROG, project no. 53615RI, 2008., 2008.
- [RD-20] Wesely, M., Parameterization of surface resistance to gaseous dry deposition regional scale numerical models. Atmos. Environ, 23:1293Ű1304, 1989.
- [RD-21] Huijnen, V., Eskes, H. J., Poupkou, A., Elbern, H., Boersma, K. F., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., Gross, A., Robertson, L., D'Isidoro, M., Kioutsioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Maurizi, A., Melas, D., Peuch, V.-H., and Zerefos, C.: Comparison of OMI NO2 tropospheric columns with an ensemble of global and European regional air quality models, Atmos. Chem. Phys., 10, 3273–3296, doi:10.5194/acp-10-3273-2010, 2010.
- [RD-22] Lansal, L. N., Martin, R. V., van Donkelaar, A., Celarier, E. A., Bucsela, E. J., Boersma, K. F., Dirksen, R., Luo, C., and Wang, Y.: Indirect validation of tropospheric nitrogen dioxide retrieved from the Ozone Monitoring Instrument: Insight into the seasonal variation of nitrogen oxides at northern midlatitudes, J. Geophys. Res., 115, D05301, doi:10.1029/2009JD012399, 2010.
- [RD-23] Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., and Wang, P.: An improved tropospheric NO2 retrieval for OMI observations in the vicinity of mountainous terrain, Atmos. Meas. Tech., 2, 401–416, doi:10.5194/amt-2-401-2009, 2009.



- [RD-24] Levelt, P.F., E. Hilsenrath, G.W. Leppelmeier, G. H. J. van Oord, P. K. Barthia, J. Tamminen, J. F. de Haan, and J. P. Veefkind. Science Objective of the Ozone Monitoring Instrument. IEEE Trans. Goe. Rem. Sens., 44, No.5:11991208, 2006a.
- [RD-25] Levelt, P.F., G.H.J. van Oord, M.R. Dobber, M. Malkki, H. Visser, J. de Vries, P. Stammes, J. Lundell, and H. Saari. The Ozone Monitoring Instrument. IEEE Trans. Goe. Rem. Sens., 44, No. 5:1093 1101, 2006b.
- [RD-26] Schoeberl, M.R., A.R. Douglass, E. Hilsenrath, P.K. Bhartia, R. Beer, J.W. Waters, M.R. Gunson, L. Froidevaux, J.C. Gille, J.J. Barnett, P.F. Levelt and P. DeCola, Overview of the EOS aura mission IEEE Transactions on Geoscience and Remote Sensing, 2006, 44, 5, 1066-1074,

- [RD-1] Dozier, J. (1981) A method for satellite identification of surface temperature fields of subpixel resolution. Rem. Sens. of Environ. 11, 221–229.
- [RD-2] Elbern, H., Strunk, A., Schmidt, H., Talagrand, O., 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion. Atmospheric Chemistry and Physics Discussions 7, 1725-1783.
- [RD-3] ESA Aerosol CCI European Space Agency Aerosol Climate Change Initiative (http://www.esaaerosol-cci.org/)
- [RD-4] Galperin M.V. (1999). Approaches for improving the numerical solution of the advection equation. In: Large-Scale Computations in Air Pollution Modelling, ed. by Z. Zlatev et al., Proc. NATO Advanced Research Workshop on Large Scale Computations in Air Pollution Modelling, Sofia, Bistritza, July 6-10, 1998. Kluiwer Academic Publishers, the Netherlands, 161-172.
- [RD-5] Galperin M.V. (2000). The Approaches to Correct Computation of Airborne Pollution Advection. In: Problems of Ecological Monitoring and Ecosystem Modelling, vol. XVII, St. Petersburg. Gidrometeoizdat, 54-68 (in Russian).
- [RD-6] Giglio, I. Kendall, J.D. (2001) Application of the dozier retrieval to wildfire characterization: a sensitivity analysis. Remote Sensing of Environment, 77(1): 34–49.
- [RD-7] Horn, H.-G., Bonka, H., Maqua, M., (1987). Measured particle bound activity size-distribution, deposition velocity, and activity concentration in rainwater after the Chernobyl accident. Journal of Aerosol Science 18, 681–684.
- [RD-8] Hyer, E. J., Reid, J. S., and Zhang, J.: An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of modis collection 5 optical depth retrievals, Atmospheric Measurement Techniques, 4, 379-408, doi:10.5194/amt-4-379-2011, 2010.
- [RD-9] Ichoku, C., Kaufman, J.Y (2005) A Method to Derive Smoke Emission Rates From MODIS Fire Radiative Energy Measurements IEEE Transactions on geoscience and remote sensing, 43, No 11
- [RD-10] Jylhä, K., (1991). Empirical scavenging coefficients of radioactive substances released from Chernobyl. Atmospheric Environment 25A (2), 263–270.
- [RD-11] Kaufman, Y. J., Justice, C. O., Flynn, L. P., Kendall, J. D., Prins, E. M., Giglio, L. Ward, D. E., Menzel, W. P., Setzer, A. W. (1998) Potential global fire monitoring from EOS-MODIS, J. Geophys. Res., 103, pp. 32215–32238
- [RD-12] Matson, M., & Dozier, J. (1981). Identification of subresolution high temperature sources using a thermal IR sensor.Photogramm. Eng. Remote Sens., 47:1311–1318, 1981
- [RD-13] Marchuk, G., 1995. Adjoint equations and analysis of complex systems. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [RD-14] Remer, L. A., Kaufman, Y. J., Tanr´e, D., Matoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kieidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, J. Atmos. Sci., 62(4), 947–973, doi:10.1175/JAS3385.1, 2005.



- [RD-15] Saarikoski, S., Sillanpää, M., Sofiev, M., Timonen, H., Saarnio, K., Teinilä, K., Karppinen, A., Kukkonen, J., Hillamo, R. (2007) Chemical composition of aerosols during a major biomass burning episode over northern Europe in spring 2006: experimental and modelling assessments. Atmosph. Environ., 41, 3577-3589.
- [RD-16] Shi Y., J. Zhang, J. S. Reid, B. Holben, E. J. Hyer, and C. Curtis, An analysis of the collection 5 MODIS over-ocean aerosol optical depth product for its implication in aerosol assimilation,
- [RD-17] Slinn, S., and Slinn, W. (1980). Predictions for particle deposition on natural waters, Atmospheric Environment, 14(9), 1013 { 1016, doi:10.1016/0004-6981(80)90032-3.
- [RD-18] Smith, F.B. and Clark, M.J. (1989). The transport and deposition of radioactive debris from the Chernobyl nuclear power plant accident with special emphasis on consequences to the United Kingdom. Meterorological Office Scientific Paper, N42, HMSO, London.
- [RD-19] Sofiev, M. (2000). A model for the evaluation of long-term airborne pollution transport at regional and continental scales. Atmospheric Environment. 34, 2481-2493.
- [RD-20] Sofiev, M. (2002). Extended resistance analogy for construction of the vertical diffusion scheme for dispersion models. J. of Geophys.Research – Atmosphere, 107, D12, doi: 10.1029/2001JD001233
- [RD-21] Sofiev, M., Galperin, M., Genikhovich, E. (2008). Construction and evaluation of Eulerian dynamic core for the air quality and emergency modelling system SILAM. NATO Science for piece and security Serties C: Environmental Security. Air pollution modelling and its application, XIX, Borrego, C., Miranda, A.I. (eds.), Springer, 699-701
- [RD-22] Sofiev M., Siljamo P., Valkama I., Ilvonen M. and Kukkonen, J. (2006). A dispersion modelling system SILAM and its evaluation against ETEX data. Atmosph.Environ., 40, 674-685.
- [RD-23] Sofiev, M., Genikhovich, E., Keronen, P., Vesala, T. (2010). Diagnosing the surface layer parameters for dispersion models within the meteorological-to-dispersion modelling interface, J. of Appl. Meteorol. and Climatology, DOI: 10.1175/2009JAMC2210.1.
- [RD-24] Sofiev, M., Soares, J., Prank, M., de Leeuw, G., Kukkonen, J. (2011) A regional-to-global model of emission and transport of sea salt particles in the atmosphere. JCR, 116, D21302, doi:10.1029/2010D014713
- [RD-25] Vira, J., Sofiev, M., 2012. On variational data assimilation for estimating the model initial conditions and emission fluxes for short-term forecasting of SOx concentrations. Atmospheric Environment 46, 318-328.
- [RD-26] Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, a, Satake, S., 2007. Adjoint inverse modeling of dust emission and transport over East Asia. Geophysical Research Letters 34, 1-6.
- [RD-27] Zhang, J. and Reid., J.S., MODIS Aerosol Product Analysis for Data Assimilation: Assessment of Level 2 Aerosol Optical Thickness Retrievals, J. Geophysical Research-Atmospheres, VOL. 111, D22207, doi:10.1029/2005JD006898, 2006.

# 2.2 Informative documents

- [ID-1] User Requirements Document (GE\_URD\_02\_00)
- [ID-2] Product Specification Document (GE\_PSD\_01\_02)



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:13 of 64

# 3 Overview of algorithms

Satellite instruments measure high resolution radiance spectra at wavelengths in the ultraviolet, visible or infrared, which provide valuable information on the chemical composition of the atmosphere for various trace gases and aerosols. From their sunsynchronous orbit at a height of ~800 km they have a global coverage within 1-5 days at a 10-50 km resolution. The homogeneity and global coverage of the satellite retrievals are ideal to estimate top-down emissions. To achieve this, the retrieved tropospheric column concentrations of a trace gas are compared with the simulated concentrations from a chemistry transport model, based on a bottom-up emission inventory. The difference between observed and modelled concentrations contains information on how to adjust the underlying trace gas emissions. This is an inverse problem which is computationally challenging because the non-local relation (sensitivity) between emission and concentration has to be found. Due to transport away from the source, life-time information of the pollutant is crucial.

Several approaches have been developed, which are applied to different time scales and emission inventory resolutions. Shorter assimilation intervals ask for fast data assimilation algorithms, while transport issues become important for high spatial resolution. The methods differ in their emission domain (global or regional), the used satellite data, and the used chemical transport model, and sensitivity derivation.

Various techniques are used: e.g. adjoint modelling, Kalman filters, ensemble Kalman filter and 4D-VAR data assimilation. Also the chemical-transport models vary: IMAGES, CHIMERE, LOTOS-EUROS, SILAM. The details of the algorithms used in the GlobEmission project are discussed in detail in section 4.



# 4 Algorithm descriptions

# 4.1 Global Emission Estimates

# 4.1.1 Introduction

The adjoint model technique coupled with the IMAGESv2 global CTM is used for the derivation of satellite-based global emission estimates. The algorithm has been fully described in the first version of the ATBD document of GlobEmission (<u>ATBD\_02\_01</u>) (Figure 1). Below we give an overview of the input data used in the source inversion exercises (Section 1.1.2), and discuss model features regarding the HCHO and SO<sub>2</sub> model simulation (Section 1.1.3). Sensitivity inversions allowing to characterize uncertainties in the emission estimates are also presented (Section 1.1.3).



Figure 1: A schematic view of the source inversion algorithm as implemented in IMAGES v2.

# 4.1.2 Input data

<u>Satellite data</u>: HCHO vertical column abundances from the OMI and GOME-2 sensors. This dataset is obtained from the UV-Vis retrieval team of BIRA-IASB and is available at http://h2co.aeronomie.be. A new dataset of  $SO_2$  column densities from the OMI instrument has been developed at BIRA-IASB and will be used to constrain the  $SO_2$  source.

Title:	Algorithm Theoretical Baseline
Issue:	3.4
Date:	31 March 2016
Page:	15 of 64

A priori emissions: Anthropogenic emissions are obtained from the RETRO 2000 database (http://retro.enes/org, Schultz et al. 2008), except over Asia where REASv2 database for year 2008 is used (Kurokawa et al., 2013). Open fires are provided by the Global Fire Emission Database (GFEDv3, van der Werf et al., 2010) and the emission factors are from the 2011 update of the database of Andreae and Merlet(2001). Vertical smoke profiles from open fires are obtained from Sofiev et al. (2013) and are based on plume top heights and MODIS fire radiative power. On average and on the global scale, half of emitted flux is injected in the boundary layer. The injection profile maps are obtained from the GlobEmission web portal (http://www.globemission.eu/data.php) and implemented in IMAGESv2. The diurnal cycle fire profile was derived based on Roberts et al.(2009). The global GFEDv3 NMVOC emission estimate is estimated at 105.4 TgVOC in 2010. Biogenic emissions of isoprene are from MEGAN-MOHYCAN inventory obtained the (http://tropo.aeronomie.be/models/isoprene.htm, Stavrakou et al., 2014). The global annual estimate of isoprene emissions is 363 Tg for 2010. The biogenic source of methanol is in Stavrakou et al. Stavrakou et al. (2011).

Anthropogenic SO<sub>2</sub> emissions are obtained from the REASv2 inventory (Kurokawa et al.2013) over Asia, from EMEP over Europe (http://www.ceip.at) and from the EDGAR3.2 FT2000 inventory for 2000 over the rest of the world. The global emissions of SO<sub>2</sub> from anthropogenic sources are estimated at 53.8 TgS in 2010. Vegetation fires are obtained from the GFEDv3 database and account for 1.3 TgS in 2010. Emissions of SO<sub>2</sub> and other sulfur compounds from continuously degassing volcanoes are constant and amount to 7.2 TgS annually (Andres and Kasgnoc, 1998). Besides direct emissions, SO<sub>2</sub> is formed through oxidation of sulfur-containing biospheric compounds like dimethyl sulfide (DMS), carbonyl sulfide (OCS), carbon disulfide (CS<sub>2</sub>), and hydrogen sulfide (H<sub>2</sub>S). The global annual photochemical source is calculated at 18.1 TgS. Dry and wet deposition account for approx. 60% of the global SO<sub>2</sub> sink, followed by oxidation by OH (20%), and heterogeneous in-cloud reactions of SO2 with H<sub>2</sub>O<sub>2</sub> (18%) and with O<sub>3</sub>. The aerosol simulation in IMAGESv2 is described in Stavrakou et al.(2013).

#### 4.1.3 HCHO simulation using the IMAGESv2 global CTM

IMAGESv2 global model provides the global distribution of about 130 chemical constituents between the Earth's surface and 22.5 km, at a resolution of 2 degrees in latitude, 2.5 degrees in longitude and 40 vertically discretized levels. Advection is driven by monthly mean operational ECMWF ERA-Interim fields, and daily fields are used for temperature, water vapour, boundary layer mixing, and cloud optical depths. As the model time step is one day, diurnal variations in the photorates and in the concentrations are accounted for through correction factors computed via a diurnal cycle simulation with a 20-minute time step. The diurnal profiles are used to estimate the formaldehyde and SO<sub>2</sub> model columns at the overpass times of the two satellites (9:30 LT for GOME-2, and 13:30 LT for OMI) from the daily averaged values calculated with a time step of one day. The modelled columns are confronted with HCHO and SO<sub>2</sub> column data (binned onto the model horizontal grid and monthly averaged), following a 4-month spin-up time. The simulated monthly averaged

 Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	16 of 64

columns account for the vertical sensitivity of the measurements through the averaging kernels, and for the sampling times of observations at each location.

The inorganic chemistry reactions, as well as the methane degradation mechanism, are as in Müller and Stavrakou (2005), with updated kinetic rates from Sander et al. (2011). The degradation mechanism for the majority of the NMVOCs is largely based on the Master Chemical Mechanism (MCM) (Saunders et al., 2003). The isoprene oxidation chemistry follows the LIMO mechanism (Peeters et al., 2014), which recycles OH more efficiently than generally assumed in models under low NOx conditions, (Peeters and Müller, 2010).

#### 4.1.3.1 HCHO yields from the oxidation of pyrogenic NMVOCs

The oxidation chemistry of pyrogenic NMVOCs is presented in Stavrakou et al.(2009). According to this study, based on the quasi-explicit Master Chemical Mechanism v3.1 (Saunders et al., 2003), we performed time-dependent simulations under high NOx conditions using the chemical solver of the KPP package (Damian et al., 2002;Sandu and Sander, 2006). Simulations start at 06:00 for a temperature of 298K at a latitude of 30 degrees in February. The model is initialized with 0.1 ppb of the considered NMVOC, 35 ppb  $O_3$ , and 100 ppb CO. The NO<sub>2</sub> concentration is kept constant and is taken equal to 1 ppb; such a choice reflects the high NOx regime associated to biomass burning events. Two HCHO yields are computed: after one day of simulation (short-term), and after 2 months (final). The short-term yield is defined as

#### $Y_{st} = (HCHO \text{ produced after 1 day})/C_0(NMVOC), (1)$

where  $C_0(NMVOC)$  is the initial concentration of the NMVOC. This yield represents the number of HCHO molecules generated by a given NMVOC one day after the injection time. The short-term yield defined in Eq. (1) provides better indication of the HCHO production that may be detected by the satellite instrument directly above biomass burning areas. The final yield is defined as

#### $Y_f = (HCHO \text{ produced})/\Delta C(NMVOC), (2)$

where  $\Delta C(NMVOC)$  is the difference between the initial and the final NMVOC concentrations. Due to the importance of both short-term and final yields in the correct representation of the HCHO production by our chemical mechanism, particular care has been taken to ensure that the IMAGESv2 calculated HCHO yields are as close as possible to the MCM yields, as is evident from Table 1.

Compounds with the highest short-term HCHO yield are ethene (1.38 mol/mol), propene (1.78 mol/mol), and 2,3- butanedione (2 mol/mol). Strongly emitted compounds like acetic acid and methanol with lifetimes of several days have very small 1-day yields. For relatively short-lived species like ethene, glycolaldehyde, propene, acetaldehyde, and isoprene, more than 80% of the final yield is reached within the first day in the box model simulations.



Table 1: Photochemical production of HCHO from pyrogenic NMVOCs. The emission estimates are 10<sup>-</sup> year averages based on GFEDv2 and the MEGAN-ECMWF inventory. Short-term and final yields are obtained from box model simulations.

	NMVOC Emission	MCM HCHO Yield (mol/mol)		IMAGESv2 HCHO Yield (mol/mol)		MCM HCHO Production (Tg/yr)	
	(Tg/yr)	Shterm	Final	Shterm	Final	Shterm	Final
Acetic acid	17.1	0.12	1	0.11	1	1.03	8.54
Methanol	8.8	0.14	1	0.14	1	1.15	8.28
Ethene	6.3	1.38	1.76	1.33	1.76	9.34	11.92
Glyoxal	6.0	0.06	0.06	0.06	0.06	0.19	0.19
Formaldehyde	5.0	1	1	1	1	5.04	5.04
Glycolaldehyde	4.8	0.80	0.87	0.76	0.88	1.92	2.10
Methylglyoxal	3.9	1	1	1	1	1.63	1.63
2,3-butanedione	3.7	2	2	2	2	2.55	2.55
Acetaldehyde	3.6	0.94	1	0.94	1	2.25	2.40
Propene	3.5	1.78	1.92	1.80	1.95	4.34	4.68
Ethane	3.2	0.03	0.99	0.03	0.99	0.09	3.21
Acetone	3.0	0.11	2	0.10	2	0.18	3.18
Propane	1.9	0.03	1.65	0.03	1.62	0.04	2.15
2-butanone	1.7	0.53	2.15	0.49	1.99	0.38	1.54
Benzene	1.7	0.05	0.31	-	-	0.03	0.20
Toluene	1.0	0.62	1.34	-	-	0.21	0.46
Others	20.8 <sup>a</sup>					5.04	7.60
Total pyrogenic	96					35	66

# 4.1.3.2 HCHO yields from the oxidation of isoprene

The LIM1 isoprene oxidation mechanism (Peeters et al.2014) is used in the current version of the model. This mechanism builds upon the previous version LIM0 (Peeters et al. 2009; Peeters and Müller, 2010; Stavrakou et al. 2010) and accounts for OH recycling in the oxidation of isoprene. To calculate the HCHO yields using LIM1, we conduct box model simulations starting at 9:00 am for a mid-latitude location in summer. Concentrations of NOx are kept constant throughout the simulation at very low NOx (40 ppt), low NOx (0.1 ppb), and high NOx conditions (1 ppb). The initial ozone concentration is set to 35 ppb. The HCHO yield from isoprene after one day of simulation is equal to 1.13, 1.54, and 2.20 mol/mol, and the final yields calculated after 2 months of simulation are equal to 1.78, 1.91, and 2.39 mol/mol, in very low, low, and high NOx regimes, respectively.

# 4.1.3.3 HCHO yields from the oxidation of anthropogenic NMVOCs

The MCM degradation mechanism for 32 anthropogenic NMVOCs is used in order to determine the HCHO yields from the NMVOCs oxidation under high (1 ppbv NO<sub>2</sub>) NO<sub>x</sub> conditions. Based on these yields and the species reactivity against OH oxidation, we derive the mean yield and mean reactivity for a lumped species OTHC (other hydrocarbons) which is implemented in IMAGESv2 (Table 2). In this way, OTHC is representative of the mixture of anthropogenic VOC compounds not explicitly included in the model. More details are given in Stavrakou et al. (in preparation).



Species	Yield
butane	2.03
2-methyl butane	2.81
pentane	2.10
hexane	2.80
2-methyl propane	2.71
heptane	2.95
1,2,4-trimethyl benzene	3.00
ethylbenzene	1.11
octane	2.98
decane	3.83
2-methylpropene	2.89

# Table 2: Anthropogenic VOCs and the corresponding HCHO yields (mol/mol) calculated in a high NOx environment.

# 4.1.4 Error analysis

Model errors might lead to potentially significant uncertainties on the derived fluxes which are, however, very difficult to estimate. To address this issue, sensitivity tests will be carried out to assess the influence of various parameters on the inferred fluxes. First, we will carry out inversions based on either GOME-2 or OMI HCHO columns, and assess the consistency between the derived emission fluxes. This will allow to investigate the robustness of the derived emissions on the choice of the satellite dataset. Furthermore, sensitivity tests where the error on the a priori emission parameters (biogenic, biomass burning) are either halved or doubled will be conducted. Another sensitivity experiment will consist in addressing the dependency of the results on the cloud filter assumed for the satellite data. For that, HCHO data with cloud cover lower than 20% will be used in the inversion, instead of 40% in the standard case study. Moreover, we have incorporated in IMAGESv2 a number of new developments in the isoprene degradation chemistry (Peeters and Müller, 2010; Stavrakou et al., 2010) which leads to HCHO formation. We will investigate the impact of these updates on the derived NMVOC flux strengths, especially over pristine forest regions, but also over mid-latitude regions in summertime. Further sensitivity studies could be conducted in order to improve the characterization of the uncertainties on the satellite-based emissions.



#### 4.1.5 References

Andreae, M. O. and Merlet, P.: Emission of trace gases and aerosols from biomass burning, Gl. Biogeochem. Cy., 15, 955–966, 2001.

Andres, R. and Kasgnoc, A.: A time-averaged inventory of subaerial volcanic sulfur emissions, J. Geophys. Res. Atmospheres, 103(D19), 25251–25261, 1998.

Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G. R.: The kinetic preprocessor KPP-a software environment for solving chemical kinetics, Computers & Chemical Engineering, 26(11), 1567–1579, 2002.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K. and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2, Atmos. Chem. Phys., 13(21), 11019–11058, 2013.

Müller, J.-F. and Stavrakou, T.: Inversion of CO and NOxemissions using the adjoint of the IMAGES model, Atmos. Chem. Phys., 5(5), 1157–1186, doi:10.5194/acp-5-1157-2005, 2005.

Peeters, J. and Müller, J.-F.: HOx radical regeneration in isoprene oxidation via peroxy radical isomerisations. II: experimental evidence and global impact, Phys. Chem. Chem. Phys., 12(42), 14227–14235, 2010.

Peeters, J., Müller, J.-F., Stavrakou, T. and Nguyen, V. S.: Hydroxyl Radical Recycling in Isoprene Oxidation Driven by Hydrogen Bonding and Hydrogen Tunneling: The Upgraded LIM1 Mechanism, The Journal of Physical Chemistry A, 2014.

Peeters, J., Nguyen, T. L. and Vereecken, L.: HOx radical regeneration in the oxidation of isoprene, Phys. Chem. Chem. Phys., 11(28), 5935–5939, 2009.

Roberts, G., Wooster, M. and Lagoudakis, E.: Annual and diurnal african biomass burning temporal dynamics, Biogeosciences, 6, 849–866, 2009.

Sander, R., Baumgaertner, A., Gromov, S., Harder, H., Jöckel, P., Kerkweg, A., Kubistin, D., Regelin, E., Riede, H., Sandu, A. and others: The atmospheric chemistry box model CAABA/MECCA-3.0, Geosc. Mod. Dev., 4, 373–380, 2011.

Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, Atmos. Chem. Phys., 6(1), 187–195, 2006.

Saunders, S., Jenkin, M., Derwent, R. and Pilling, M.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmos. Chem. Phys., 3(1), 161–180, 2003.



Schultz, M. G., L. Backman, Y. Balkanski and al., et: REanalysis of the TROpospheric chemical composition over the past 40 years (RETRO): A long-term global modeling study of tropospheric chemistry, 48/2007 report on Earth System Science of the Max Planck Institute for Meteorology, edited by M. P. I. for Meteorology, Hamburg, Jülich, Germany., 2007.

Sofiev, M., Soares, J., Hakkarainen, J., Ermakova, T. and Vankevich, R.: A global wildfire emission and atmospheric composition impact in 2000-2012, in EGU General Assembly Conference Abstracts, vol. 15, p. 12854., 2013.

Stavrakou, T., J. Peeters, and J.-F. Müller: Improved global modelling of HOx recycling in isoprene oxidation: evaluation against the GABRIEL and INTEX-A aircraft campaign measurements, *Atmos. Chem. Phys.*, 10, 9863-9878, 2010.

Stavrakou, T., Guenther, A., Razavi, A., Clarisse, L., Clerbaux, C., Coheur, P.-F., Hurtmans, D., Karagulian, F., De Mazière, M., Vigouroux, C., Amelynck, C., Schoon, N., Laffineur, Q., Heinesch, B., Aubinet, M., Rinsland, C. and Müller, J.-F.: First space-based derivation of the global atmospheric methanol emission fluxes, Atmos. Chem. Phys., 11(10), 4873–4898, doi:10.5194/acp-11-4873-2011, 2011.

Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Van Roozendael, M., Guenther, A., Wild, M. and Xia, X.: Isoprene emissions over Asia 1979 - 2012: impact of climate and land-use changes, Atmos. Chem. Phys., 14(9), 4587–4605, doi:10.5194/acp-14-4587-2014, 2014.

Stavrakou, T., Müller, J.-F., Boersma, K., Kurokawa, J., Ohara, T., Zhang, Q.: Key chemical NO x sink uncertainties and how they influence top-down emissions of nitrogen oxides, Atmos. Chem. Phys., 13, 9057-9082, 2013.

Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., Werf, G. R. van der, Giglio, L. and Guenther, A.: Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns, Atmos. Chem. Phys., 9(3), 1037–1060, doi:10.5194/acp-9-1037-2009, 2009.

Stavrakou, T., Peeters, J. and Müller, J.-F.: Improved global modelling of HO\_x recycling in isoprene oxidation: evaluation against the GABRIEL and INTEX-A aircraft campaign measurements, Atmos. Chem. Phys., 10(20), 9863–9878, doi:10.5194/acp-10-9863-2010, 2010.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y. and Leeuwen, T. T. van: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009), Atmos. Chem. Phys., 10(23), 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:21 of 64

# 4.2 Regional Emission Estimates

The DECSO algorithm (Daily Emission estimation Constrained by Satellite Observations, Mijling and Van der A, 2012) has been applied successfully for NOx emission estimation in East Asia, Middle East, South Africa, and India. In the following sections the algorithm is explained in detail, ending with an overview of the different algorithm versions.

# 4.2.1 Input data

The tropospheric columns of NO2 and SO2 of the satellite instruments OMI (on Aura) and GOME-2 (on Metop) are used. For both instruments, the tropospheric NO<sub>2</sub> columns are taken from the KNMI retrieval product (DOMINO version 2), described by *Boersma et al.* (2011) and available through the TEMIS portal (http://www.temis.nl). SO<sub>2</sub> data is obtained from NASA/Goddard Space Flight Center (http://so2.gsfc.nasa.gov/).

# 4.2.2 Algorithm description

Below, we derive the algorithm for NOx estimation estimates, but it can be applied to other short-lived trace gases as SO2 and CH2O. Because of the rapid cycling between NO and NO<sub>2</sub>, and the shift towards NO<sub>2</sub> during night time due to lack of photolysis, our analysis is based on the bookkeeping of NO<sub>x</sub> rather than NO<sub>2</sub>, which is the observed quantity. We consider a time interval t=[0,T] between two data assimilation moments, in our case the 24 h period between two overpasses of the satellite instrument. An schematic overview of the algorithm is depicted in Figure 3.



Figure 1: Flow chart of the DECS O algorithm. Central is the CHIMERE model, which is approximated by the simplified 2D transport model to get sensitivity information used in the inversion by the Kalman filter.

At the core of the algorithm is the chemical transport model (CTM) which calculates the concentration fields of NO and NO<sub>2</sub> at t=T from the initial fields at t=0, given the meteorological conditions and a certain emission database. After the model run, however, there is no information available of the sensitivity of the final NO<sub>x</sub> column concentration on the emission field; information needed by the Kalman filter for the emission inversion. Instead, we will derive a simplified 2D transport equation which approximates this relation analytically. It uses trajectory analysis to transport NO<sub>x</sub> columns over the model domain. The aging of the NO<sub>x</sub> column is described by an effective lifetime, chosen in such way that it

	Title: Ref:	Algorithm Theoretical Baseline GF ATBD
	Issue:	3.4
GLOBEMISSION	Date: Page:	31 March 2016 22 of 64

minimizes the calculated concentrations and the simulated concentrations by the CTM at t=T. The concentration fields from the CTM are also used to construct the forecasted tropospheric NO<sub>2</sub> column concentration, by interpolation towards the satellite footprint, extension with a climatological column to the tropopause, and application of the averaging kernel from the satellite retrieval method.

#### A priori emission inventory

For the first DECSO emission estimates of NOx for the East China region, we use the INTEX-B emission inventory by *Zhang et al.* (2009, and private communication). In later algorithm versions and other regions, other emission inventories are used (see Section 4.2.4). INTEX-B covers Asia on a  $0.2^{\circ} \times 0.2^{\circ}$  resolution, containing the yearly totals of SO<sub>2</sub>, NO<sub>x</sub>, CO, VOC, PM<sub>10</sub>, PM<sub>2.5</sub>, BC, and OC by four sectors (power, industry, residential, and transportation) for the year 2006. The NO<sub>x</sub> emission totals are interpolated to the model grid and distributes over three chemical components: NO (90% of the NO equivalent mass of NO<sub>x</sub>), NO<sub>2</sub> (9.2%), and HONO (0.8%). The emissions are disaggregated to hourly values using sector-specific weekly and diurnal factors; no monthly cycle is postulated.

#### Chemical transport model

The CHIMERE multi-scale model (*Schmidt* et al., 2001; *Bessagnet* et al., 2004) is primarily designed to produce daily forecasts of ozone, aerosols and other pollutants and make long-term simulations for emission control scenarios. In the presented configuration, CHIMERE has been implemented over East Asia (18°N to 50°N and 102°E to 132°E), simulating the atmosphere in 8 layers up to 500 hPa, with a horizontal resolution of  $0.25^{\circ}\times0.25^{\circ}$ . The meteorological data is taken from the deterministic forecast of the European Centre for Medium-Range Weather Forecasts (ECMWF), which is given on 91 atmospheric layers for a horizontal resolution of approximately  $25\times25$  km<sup>2</sup>. The boundary conditions for the domain border and top are taken from monthly climatologies (no nested run is performed). To reduce the effect of boundary values on the region of interest, the domain boundaries have been chosen over relatively unpopulated areas.

In Europe, CHIMERE has been extensively intercompared to other urban air quality models (e.g. *Vautard et al.*, 2006) and evaluated against ground-based measurements and satellite data (e.g. *Blond et al.*, 2007). For China, validation results can be found in *Mijling et al.*, 2009.

#### Derivation of the simplified 2D transport equation

Let **e**, **c**(0), and **c**(*T*) be two-dimensional fields, representing the time-averaged emission, the NO<sub>x</sub> column concentration at t=0, and at t=T, respectively. They are written as vectors in an *n*-dimensional space, where *n* represents the total number of grid cells in the model domain. In our case, the dimension of state space *n* is  $121 \times 129 = 15,609$ .

For a certain grid cell at t=T, the NO<sub>x</sub> column is composed of an aged column from t=0 which has been transported into this grid cell, and a column of NO<sub>x</sub> which has been emitted



during [0,T], aged, and transported into this grid cell. This can be written by the matrix equation:

$$\mathbf{c}(T) = \mathbf{G}\mathbf{c}(0) + \mathbf{H}\mathbf{e} \tag{1}$$

Matrix **G** describes the advection and decay of the initial concentration  $\mathbf{c}(0)$  over the model grid. Matrix **H** describes how the concentrations at t=T change if the emissions in time interval *T* change, i.e. its matrix elements represent the sensitivities of the NO<sub>x</sub> column concentrations to the NO<sub>x</sub> emission at the model grid:

$$H_{ij} = \frac{\partial c_i}{\partial e_j} \tag{2}$$

Both **G** and **H** have size  $n \times n$ , but are sparse if transport to other grid cells covers only a small fraction of the model domain, which is true for short-lived species such as NO<sub>x</sub> (~8h) and typical wind speeds (~24 km/h at 4 Beaufort) in our mesoscopic domain.

The first term in equation (1) describes the advection and decay of the initial concentration over the model grid. The concentration of  $NO_x$  particles in a certain grid cell column *i* at t=T can then be written as

$$c_i^{\rm G}(T) = \sum_j e^{-k_j T} \frac{a_j}{a_i} \Omega_{ij}(T) c_j(0) \quad \text{, with} \quad k_j = 1/\tau_j$$
(3)

in which the transport kernel  $\Omega_{ij}$  (*T*) represents the fraction of the area of cell *j* which is transported to cell *i* during time interval *T*. To ensure mass conservation, the equation is scaled with the area *a* of the corresponding grid cells. For a regular grid over latitude and longitude *a* will depend on latitude  $\varphi$ :  $a_j/a_i = \cos(\varphi_j)/\cos(\varphi_i)$ . The number of NO<sub>x</sub> particles in the plume will decay exponentially because of aging, which is described by an effective lifetime  $\tau_j$  (or the reciprocal lifetime  $k_j$ ) from its source location.

The second term in equation (1) represents all "fresh"  $NO_x$  which has been emitted at a certain moment in the time interval and subsequently transported during the remaining time. Analogous to equation (3), and integrating over all time dependent emission contributions e(t) during *T*, we derive the expression:

$$c_{i}^{\rm H}(T) = \sum_{j} \frac{a_{j}}{a_{i}} \int_{0}^{T} e^{-k_{j}t} \Omega_{ij}(t) e_{j}(T-t) dt$$
<sup>(4)</sup>

The inversion only adjusts time-averaged emissions; the diurnal and weekly cycle of the emissions are described by an emission modulation function f(t), which is copied from the chemistry transport model:

$$e_{j}(t) = f_{j}(t)e_{j} \quad , \quad \left\langle f(t) \right\rangle_{1 \text{ week}} = 1$$
(5)

Substituted in equation (4) this results in:



$$c_i^{\mathrm{H}}(T) = \sum_j \frac{a_j}{a_i} \left( \int_0^T e^{-k_j t} \Omega_{j \to i}(t) f_j(T-t) dt \right) e_j$$
(6)

Relating equation (4) and (6) to equation (1) gives us expressions for the elements of matrix **G** and **H** in terms of the (unknown) reciprocal lifetimes  $k_j$  and the (known) transport kernel  $\Omega(t)$ :

$$G_{ij}(k_j) = \frac{a_j}{a_i} e^{-k_j T} \Omega_{ij}(T)$$
<sup>(7)</sup>

$$H_{ij}(k_{j}) = \frac{a_{j}}{a_{i}} \int_{0}^{T} e^{-k_{j}t} \Omega_{ij}(t) f_{j}(T-t) dt$$
(8)

We will call matrix equation (1) with matrix elements defined in (7) and (8) the *simplified 2D transport equation*. Because it does not contain boundary conditions, it does not account for NO<sub>x</sub> concentrations which are transported from the outside into the model domain. This will affect sensitivity relations  $\partial c_i/\partial e_j$  (and consequently the assimilation results) close to the domain border. We suppress this boundary effect by choosing our domain borders (where possible) in remote areas with low emissions.

Equation (8) describes the sensitivity of the  $NO_x$  concentration in cell *i* on the  $NO_x$  emission in cell *j*. However, we are interested in the sensitivity of the  $NO_2$  concentration on the  $NO_x$  emission:

$$H_{ij}^* = \frac{\partial c_i^{NO2}}{\partial e_j^{NOx}}$$
(9)

From the CTM simulation we obtain the NO<sub>2</sub>/NO<sub>x</sub> ratio  $\gamma_i$  for all cells *i* at the moment of assimilation *T*:

$$c_i^{NO2}(T) = \gamma_i c_i^{NOx}(T) \quad , \quad 0 \le \gamma_i \le 1$$
(10)

From this and equation (8) we get the sensitivities for  $NO_2$  column concentrations to  $NO_x$  emissions, both at the model grid:

$$H_{ij}^{*}(k_{j}) = \gamma_{i}H_{ij} = \gamma_{i}\frac{a_{j}}{a_{i}}\int_{0}^{T}e^{-k_{j}t}\Omega_{ij}(t)f_{j}(T-t)dt$$
(11)

#### The transport kernel

Driving force behind the transport of NO<sub>x</sub> is the time and space dependent wind field  $\mathbf{u}(x,y,z,t)$ . In the simplified 2D transport equation the transport is described by matrix function  $\Omega(t)$ , whose column *j* can be interpreted as the advection of the tropospheric column of NO<sub>x</sub>

Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	25 of 64

of grid cell *j* over the model grid during the time interval [T-t,T]. We discretize  $\Omega(t)$  at  $n_T$  instants  $t_n$ , equally distributed over the assimilation time interval:

$$\Delta t = T / n_T \quad , \quad t_n = \left( n - \frac{1}{2} \right) \Delta t \quad , \quad n = 1, \dots, n_T$$
(12)

If  $n_T$  is sufficiently large such that the transport distance is small during  $\Delta t$  (at wind speed v and grid cell size  $\Delta x$  this implies  $\Delta t \ll \Delta x/v$  must hold), the integral in equation (8) can be approximated by a summation over sparse matrix operations

$$H_{ij}(k_j) \approx \frac{a_j}{a_i} \sum_{n=1}^{n_T} e^{-k_j t_n} \Omega_{ij}(t_n) f_j(T - t_n) \Delta t$$
<sup>(13)</sup>

The elements of  $\Omega_{ij}$  are calculated for each  $t_n$  by trajectory analysis, which must take the divergence of the wind field and the height dependence of wind and NO<sub>x</sub> concentration into account. In remote areas, for example, the bulk of the NO<sub>x</sub> tends to be located higher up in the troposphere (mainly because of uplifting and washing out of the NO<sub>x</sub> in the lower troposphere) where wind flows generally faster. A practical method for fast calculation is by dividing the source grid cell j in a large amount of sub grid cells  $n_S$  and start the trajectory for each sub cell at a unique height, distributed according to the particle density in the NO<sub>x</sub> column. Each sub cell is transported along the time dependent, two-dimensional wind field vertically interpolated to this height. Then  $\Omega_{ij}$  (t) can be found by counting the number of sub cells which originate in cell j at time T-t and are within cell i at time T.

#### The effective lifetime

The variation in lifetime, however, is too large to simply assume a fixed value for  $\tau$ . The lifetime of NO<sub>x</sub> can vary between several hours and several days, depending on factors such as temperature, sunlight, precipitation, altitude, and presence of other chemical species and aerosols. This dependence is too intricate to make a straightforward parameterization of  $\tau$  which could be used in look-up tables. Instead, we use the results of the forward chemical transport model run to retrieve information on the NO<sub>x</sub> lifetime.

The matrix elements of **G** and **H** in equations (7) and (8) depend on the unknown reciprocal lifetimes **k** associated with the grid cells. The residue **r** is defined as the difference between the column concentrations  $\mathbf{c}^{\text{CTM}}$  calculated with the CTM and the concentrations calculated with the simplified 2D transport equation, at t=T:

$$\mathbf{r}(\mathbf{k}) = \mathbf{c}^{\text{CTM}}(T) - \left(\mathbf{G}(\mathbf{k})\mathbf{c}^{\text{CTM}}(0) + \mathbf{H}(\mathbf{k})\mathbf{e}\right)$$
(14)

The effective lifetime field is found by minimizing the residue for **k** numerically. The exponential dependence in k of the matrix elements, and the reciprocal dependence of  $\tau$  of k, makes the residue function relatively insensitive for variations in both low and high values of  $\tau$ . Unrealistic high or low lifetime values are corrected by constraining the lifetime  $\tau$  between 2 h and 48 h. Far from emission sources, the lifetime can only be determined from the decay of the small background field, which contributes very weakly to the residue function. A good a priori lifetime field is therefore essential, and is taken from the results of the previous day,



assuming that day-to-day variations are small.

#### Comparing model simulations with satellite observations

For NO<sub>2</sub>, not all satellite retrievals are included in the data assimilation. Retrievals at cloudy conditions (cloud fractions larger than 20%) are filtered out, to reduce the influence of the modelled NO<sub>2</sub> column below the clouds in the retrieval. For bright surfaces such as snow (surface albedo larger than 20%) the cloud fraction from the cloud algorithm becomes uncertain, hence these retrievals are also discarded. Retrievals with clouds below 800 hPa are filtered out because the possible intersection of the cloud with the NO<sub>x</sub> bulk makes the retrieval too sensitive for the exact cloud height. For OMI, the large outer 4 pixels at either side of the swath are filtered out, as are pixels affected by the row anomalies (which appear since June 2007). For GOME-2 we discard the backscan pixels (which are too large), and retrievals in the zoom-mode of the instrument (which are of unknown quality).

The CHIMERE model calculates daily tropospheric columns up to 500 hPa. In order to extend the NO<sub>2</sub> profiles from the model ceiling to the tropopause, we add a climatological partial column for this part of the free troposphere. This climatology was compiled from a 2003–2008 run of the global chemistry transport model TM5, described in *Huijnen et al.* (2010), at a  $2^{\circ}\times3^{\circ}$  resolution and 34 atmospheric layers. Although in populated areas the added free tropospheric column contributes only a few percent to the total tropospheric column, it can account for up to 50% in remote areas, where the tropospheric NO<sub>2</sub> column is small.

We construct a representative vertical NO<sub>2</sub> profile  $\mathbf{p}_i$  for the footprint area by taking a weighted average of the modeled vertical profiles  $\mathbf{p}_j^c$  in contributing grid cells, taking weight factor  $W_{ij}$  from the surface fraction of grid cell *j* which is covering footprint *i*:

$$\mathbf{p}_{i} = \sum_{j} W_{ij} \mathbf{p}_{j}^{c} = \mathbf{W} \mathbf{p}^{c} \text{, with } W_{ij} \text{ normalized} \sum_{j} W_{ij} = 1$$
(15)

Afterwards, the interpolated profile  $\mathbf{p}_i$  is rebinned to the model layers used in the tropospheric NO<sub>2</sub> retrieval algorithm, so the averaging kernel  $\mathbf{A}_i$  can be applied to get the modeled, corrected NO<sub>2</sub> column at satellite footprint *i* (which is the observable quantity):

$$y_i = \mathbf{A}_i \mathbf{p}_i \tag{16}$$

For the inversion we will also need the Jacobian of the model, i.e. the matrix  $\mathbf{H}$  which represents the linearization of the model around a certain emission field  $\mathbf{e}$ :

$$\Delta \mathbf{y} = \mathbf{H} \Delta \mathbf{e} \tag{17}$$

**H** describes the sensitivity of the modeled observations **y** (in observation space) to changes in model emissions **e** (in state space). We want to find an expression of **H** in terms of the model sensitivities we found in equation (11). The gridded concentrations are interpolated to observation footprints by applying the interpolation matrix **W** from equation (15):

$$\Delta \tilde{\mathbf{y}} = \mathbf{W} \mathbf{H}^* \Delta \mathbf{e} \tag{18}$$

R	Title: Ref.: Issue:	Algorithm Theoretical Baseline GE_ATBD 3.4
	Date:	31 March 2016
	Page:	27 of 64

in which  $\Delta \tilde{\mathbf{y}}$  represents the change of the total tropospheric NO<sub>2</sub> columns in observation space, simulated by our model. Note that there is no information on the vertical profiles, which complicates a direct application of the averaging kernels. Therefore we assume that an emission change  $\Delta \mathbf{e}$  does change the total column value  $\tilde{y}_i$  for modeled observation *i*, but does not change its vertical profile shape, so we can write the change in averaged, modeled observation in terms of the change of modeled observations:

$$\Delta y_i = \alpha_i \Delta \tilde{y}_i \quad , \quad \alpha_i \equiv \left(\frac{\mathbf{A}_i \mathbf{p}_i}{y_i^*}\right) \tag{19}$$

Combining equations (18) and (19), we get the expression for the Jacobian, in terms of the model sensitivities, the interpolation matrix, and the averaging kernels:

$$\Delta y_i = \alpha_i \sum_j \sum_k w_{ij} H_{jk} \Delta e_k \implies \mathbf{H} = \operatorname{diag}(\boldsymbol{\alpha}) \mathbf{W} \mathbf{H}^*$$
(20)

in which  $\mathbf{H}^*$  is the  $(n \times n)$  model sensitivity matrix described by equation (11) for NO<sub>2</sub> columns to NO<sub>x</sub> emissions on the model grid, and **H** is the  $(m \times n)$  sensitivity matrix for NO<sub>2</sub> column observations to gridded NO<sub>x</sub> emissions, to be used in the Kalman equations below.

#### The Kalman filter for emission estimation

The relation between column concentration difference vector  $\Delta \mathbf{y}$  and emission update vector  $\Delta \mathbf{e}$  can be written as  $\Delta \mathbf{y} = \mathbf{H}\Delta \mathbf{e}$ . Note that solving  $\Delta \mathbf{e}$  from  $\Delta \mathbf{y}$  and  $\mathbf{H}$  is an underdetermined problem. Furthermore, the errors in  $\mathbf{y}$  are large, and they would propagate non-locally in the solution, causing strong fluctuations in the emissions if this assimilation scheme is applied iteratively. To deal with these issues we use the Kalman filter, which calculates for every assimilation step the most probable emission field and its covariance, taking into account the errors in the modeled emissions and representation, and the errors in the observed concentrations. Due to the non-linearity of the problem the extended Kalman filter is used, which linearizes about the current mean and covariance:

State vector forecast 
$$\mathbf{e}^{f}(t_{i+1}) = M_{i} [\mathbf{e}^{a}(t_{i})]$$
 (K1)

Error covariance forecast 
$$\mathbf{P}^{f}(t_{i+1}) = \mathbf{M}_{i}\mathbf{P}^{a}(t_{i})\mathbf{M}_{i}^{T} + \mathbf{Q}(t_{i})$$
 (K2)

Kalman gain matrix 
$$\mathbf{K}_{i} = \mathbf{P}^{f}(t_{i})\mathbf{H}_{i}^{T}[\mathbf{H}_{i}\mathbf{P}^{f}(t_{i})\mathbf{H}_{i}^{T} + \mathbf{R}_{i}]^{-1}$$
 (KG)

State vector analysis 
$$\mathbf{e}^{a}(t_{i}) = \mathbf{e}^{t}(t_{i}) + \mathbf{K}_{i}(\mathbf{y}_{i} - H_{i} [\mathbf{e}^{t}(t_{i})])$$
 (K3)

Error covariance analysis 
$$\mathbf{P}^{a}(t_{i}) = (\mathbf{I} - \mathbf{K}_{i}\mathbf{H}_{i}) \mathbf{P}^{f}(t_{i})$$
 (K4)

The interpretation of the quantities is as follows:

- $\boldsymbol{e}^a, \boldsymbol{e}^f$  analysis and forecast of the  $NO_x$  emissions.
- $\mathbf{P}^{f}$ ,  $\mathbf{P}^{a}$  error covariance matrices  $(n \times n)$  of the forecasted emissions  $\mathbf{e}^{f}$  and the analysis of the emissions  $\mathbf{e}^{a}$ .



- *M* model describing temporal evolution of the emissions. *M* applied on the true state is assumed to introduce no bias:  $\mathbf{e}^{t}(t_{i+1}) = M_{i} [\mathbf{e}^{t}(t_{i})] + \varepsilon_{i}$ , in which the noise is normally distributed around 0 with covariance **Q**.
- **Q** covariance matrix  $(n \times n)$  of the modeled emissions (see Section 0).
- **M**  $(n \times n)$  matrix representation of emission model *M*.
- **y** observations of tropospheric  $NO_2$  columns.
- *H* observation operator which relates the emissions of NO<sub>x</sub> in the model grid to the observable tropospheric column concentrations of NO<sub>2</sub>. *H* applied to the true state is assumed to introduce no bias:  $\mathbf{y}_i^{o} = H_i [\mathbf{e}^t(t_i)] + \boldsymbol{\varepsilon}_i$ , in which the noise is normally distributed around 0 with covariance **R**.
- **R**  $(m \times m)$  matrix describing the covariances of the observation operator *H* (see Section 0).
- **H**  $(m \times n)$  Jacobian of the observation model *H*, linearized around state  $\mathbf{e} = \mathbf{e}^a$ :  $\mathbf{y}^f = \mathbf{y}^a + \mathbf{H}(\mathbf{e}^f - \mathbf{e}^a)$ , describing how the tropospheric NO<sub>2</sub> column of observation *i* changes when the NO<sub>x</sub> emission  $\varepsilon$  in grid cell *j* is changed.

#### Covariance of the observation minus forecast

The covariance of the observation minus forecast of the column concentration consists of three independent components:

$$\boldsymbol{\Sigma}_{\text{OmF}} = \boldsymbol{\Sigma}_{\text{obs}} + \boldsymbol{\Sigma}_{\text{repr}} + \mathbf{H}\mathbf{P}^{\text{f}}\mathbf{H}^{\text{T}} = \mathbf{R} + \mathbf{H}\mathbf{P}^{\text{f}}\mathbf{H}^{\text{T}}$$
(21)

The observation error of the tropospheric  $NO_2$  column is composed of errors by the measurement noise of the satellite instrument and the spectral fitting, errors related to the separation of the troposphere and the stratosphere, and errors due to retrieval method parameters, such as clouds, surface albedo and a priori profile shape (*Boersma et al.*, 2004). The second part describes representation error originating from an inaccurate CTM (due to errors in e.g. meteorology or chemistry scheme), and errors introduced by adding a climatological free tropospheric column, and interpolating grid values to the satellite footprint. Together with the observation errors they are contained in covariance matrix **R**. The last part of the OmF error describes how errors in the emission estimation propagate into the simulated column concentrations. The sensitivity matrix **H** is assumed to be exactly known; the error made by the approximating **H** with the simplified 2D transport equation is added to **R**.

It is the balance between **R** and  $\mathbf{HP}^{f}\mathbf{H}^{T}$  which determines how much information from the observed concentration difference is used to update the emission estimates. We work out a practical method to estimate the covariance matrix **R**. By neglecting spatial correlations the OmF error  $\sigma_{OmF}$  for observation *i* can be written, analogous to equation (21) as

$$\sigma_{\text{OmF},i}^2 = \sigma_{\text{obs},i}^2 + \sigma_{\text{repr},i}^2 + \sigma_{\text{prem},i}^2 = \sigma_{\text{R},i}^2 = + \sigma_{\text{prem},i}^2$$
(22)

 Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	29 of 64

in which  $\sigma^2_{\text{prem},i}$  is the propagated emission variance of matrix  $\mathbf{HP}^{f}\mathbf{H}^{T}$ . The observation error  $\sigma_{\text{obs}}$  is known from the satellite product: typically they have a dominating absolute error at low values (around 0.5  $10^{15}$  molecules/cm<sup>2</sup>), and a dominating relative error at high values (around 30–45%). The representation error  $\sigma_{\text{repr}}$  is unknown, but is assumed to be relative to the simulated tropospheric column concentration *y* with a fixed  $\varepsilon_{\text{rel}}$ :

$$\sigma_{\operatorname{repr},i} = \varepsilon_{\operatorname{rel}} y_i \tag{23}$$

For each assimilation a large quantity of observations and forecasts are available, enabling the calculation of  $\sigma_{repr}$  from OmF statistics, using the reduced  $\chi^2$  criterion:

$$\chi^{2}_{\rm red} = \frac{1}{m} \sum_{i} \left( \frac{y_{\rm obs,i} - y_{i}}{\sigma_{\rm OmF,i}} \right)^{2}$$
(24)

We calculate  $\varepsilon_{rel}$  such that  $\chi^2_{red}=1$ , meaning that the variation in OmF is well described by its error  $\sigma_{OmF}$ : the distribution of  $(y_{obs}-y)/\sigma_{OmF}$  will be Gaussian around 0 with standard deviation 1. By substituting (22) and (23) in (24) we find the equation:

$$f(\varepsilon_{\rm rel}) = \sum_{i=1}^{m} \frac{(y_{\rm obs,i} - y_i)^2}{\varepsilon_{\rm rel}^2 y_i^2 + \sigma_{\rm obs,i}^2 + \sigma_{\rm prem,i}^2} - m = 0$$
(25)

We can find the root of *f* numerically using its derivative to  $\varepsilon_{rel}$  in Newton's method (e.g. for OMI observations in May 2008,  $\varepsilon_{rel}=33\%$ ). With the daily estimated value of  $\varepsilon_{rel}$  we calculate  $\sigma_{R,i}$  according to (22). From these errors we construct the covariances by decomposing the error covariance matrix as

$$\mathbf{R} = \operatorname{diag}(\boldsymbol{\sigma}_{\mathrm{R}}) \, \mathbf{C} \, \operatorname{diag}(\boldsymbol{\sigma}_{\mathrm{R}}) \tag{26}$$

in which **C** is a correlation matrix which elements only depend on the distance between two observation footprints:  $C_{ij} = g(r_{ij})$ . We model g exponentially dependent on distance:

$$g(r_{ij}) = \exp(-r_{ij}/L) \tag{27}$$

in which L is the correlation length. Best inversion results are obtained by taking L small with respect to the footprint size (we will use L=10 km); in this case the improved condition of the matrix which is inverted in (KG) suppresses spatially oscillating solutions. Correlations between footprints at a distance larger than 6L are considered insignificantly: corresponding matrix elements  $C_{ij}$  are set to zero, resulting in a sparse covariance matrix **R**.

#### Emission covariances and inversion behavior

For the Kalman filter approach we need to assess the emission model  $\mathbf{M}$ , its error covariance  $\mathbf{Q}$ . For these matrices it is important to find realistic estimates, which will optimize the assimilation for both convergence speed and noise reduction.

Anthropogenic emission trends, if present, are usually very gradual, justifying a persistent

 Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	30 of 64

emission model, stating that tomorrow's emissions are equal to today's emissions. This implies that the emission model reduces to the identity matrix: M=I. Note that our algorithm adjusts averaged emission totals; the diurnal and weekly cycle is modulated already by the CTM, see equation (5).

To assess the influence of the emission covariance  $\mathbf{Q}$  on the inversion behavior we analyze the Kalman equations for the simple case where the concentration y only depends on one emission source x with a constant sensitivity factor h. The Kalman filter reduces to the following scalar equations, and matrix  $\mathbf{Q}$  becomes a scalar quantity q, dictating how much the error of the emission increases between two assimilation moments. We find an expression for the evolution of the error analysis  $\sigma_i^{a}$  in terms of its predecessor  $\sigma_{i-1}^{a}$ :

$$\sigma_i^{a^2} = \frac{\sigma_{R,i}^2}{h^2 \left( \sigma_{i-1}^{a^2} + q^2 \right) + \sigma_{R,i}^2} \left( \sigma_{i-1}^{a^2} + q^2 \right)$$
(28)

We can see with this formula that the limit value of the assimilation error  $\sigma^{a}$  depends on the sensitivity h (a higher h results in faster convergence to a lower  $\sigma^{a}$ ), the observation and representation error  $\sigma_{R}$  (a lower  $\sigma_{R}$  results in a lower  $\sigma^{a}$ ), and emission error increment q. q is the only unknown parameter and its value will affect the noise and convergence in the assimilation. A high emission uncertainty q results in a large noise on the assimilated emission, and a low uncertainty q leads to slow convergence.

We model q with a dominating absolute error  $\varepsilon_{abs}$  at low emissions, shifting to a dominating relative error  $\varepsilon_{rel}$  at high emissions:

$$q(x) = \varepsilon_{abs} \exp(-\varepsilon_{rel} x / \varepsilon_{abs}) + \varepsilon_{rel} x$$
(29)

Note that the absolute error component is essential to be able to pick up changing emissions in areas where zero emission are defined. For our assimilation setup we construct matrix  $\mathbf{Q}$  analogous to (26) and (27), modeling its errors according to (29) with  $\varepsilon_{abs}=0.02 \ 10^{15}$  molecules/cm<sup>2</sup>/h and  $\varepsilon_{rel}=5\%$ , and allowing for weak covariances with its nearest neighbors by taking the covariance length L=10 km. These values are a trade-off between convergence speed and noise reduction, without loosing the possibility to pick up new emission sources.

#### The Kalman gain

After establishing the covariance matrices as described in the above sections, the Kalman equations can be solved numerically. Calculation time and storage space is reduced by making optimal use of the sparseness of the matrices. Point of concern is the inversion of the symmetric matrix  $\mathbf{A} = \mathbf{H}\mathbf{P}^{f}\mathbf{H}^{T} + \mathbf{R}$  in the Kalman Gain (equation (KG)), which generally is ill-conditioned and contains a null-space. Note that  $\mathbf{A}$  being symmetric and positive semi-definite implies that  $\mathbf{A}$  has real, positive eigenvalues. The eigenvalues spectrum of matrix  $\mathbf{A}$  consists of only a few large eigenvalues and many smaller ones, which makes solving the inverse of  $\mathbf{A}$  very sensitive to noise. This sensitivity issue can be avoided by approximating  $\mathbf{A}$  with a decomposition

$$\mathbf{A} \approx \mathbf{U} \wedge \mathbf{U}^{\mathrm{T}} \implies \mathbf{A}^{-1} \approx \mathbf{U} \wedge^{-1} \mathbf{U}^{\mathrm{T}}$$
(30)

Title: Ref.: Issue:	Algorithm Theoretical Baseline GE_ATBD 3 4
Date: Page:	31 March 2016 31 of 64

in which  $\Lambda$  is a diagonal matrix consisting of the  $m_r$  largest eigenvalues of  $\Lambda$ . Here we use the numerical software library ARPACK (*Lehoucq et al.*, 1998) for a fast calculation of the largest eigenvalues and corresponding eigenvectors. We take  $m_r$  such that the sum of the largest eigenvalues account for 98% of the value of the trace of  $\Lambda$ , with a practical maximum of 1800 eigenvalues. As a result the condition number (the ratio of the largest eigenvector) is reduced to an order of 1000. Note that from DECSO version 2 onward, the approximate inversion of the Kalman Gain matrix is replaced by a full matrix inversion by LDL decomposition.

#### Calculation of the analysis covariance

Correlations in the emission analysis are introduced because different emission grid cells contribute to an observation in the satellite footprint. Although the covariances can be strongly non-local (e.g. when fast winds transport long living NO<sub>x</sub> over remote areas), normally the covarying emission area is localized within a certain distance of the grid cell. Therefore it is not necessary to calculate all  $n \times n$  elements of the analysis covariance matrix  $\mathbf{P}^{a}$ . Instead we calculate the diagonal elements  $P^{a}_{ii} = \sigma_{i}^{2}$  using (K4). Off-diagonal, we only calculate the covariance for grid cells within a radius of 300 km. Only correlations larger than 0.01 are supposed to contribute significantly to the analysis:

$$\left|\rho_{ij}\right| = \left|\frac{P_{ij}^{a}}{\sigma_{i}\sigma_{j}}\right| > 0.01 \tag{31}$$

By neglecting all smaller correlations the analysis covariance matrix becomes a sparse matrix.

#### Emission inventory update

After calculating new emissions we have to make certain assumptions to use this data to update the emission database. At this time we only adjust the  $NO_x$  emissions, neglecting the possibility that a change in anthropogenic  $NO_x$  is related to a change in other anthropogenic emissions such as particulate matter and CO. Furthermore, we assume that the emission change is due to anthropogenic sources. Since we do not know how the new  $NO_x$  emissions can be attributed to the different sectors (e.g. power, industry, residential, and transportation in the INTEX-B inventory for the Chinese region), we scale emissions in all sectors by ratio, assuming that the introduced error (through different sectorial diurnal and weekly emission cycles) is small. For new emission sources, where no sector information is known, the new  $NO_x$  emissions are distributed evenly over all sectors. Finally, information about the injection height of the new emission cannot be inferred with the DESCO algorithm. Instead, emissions at all heights will be scaled by ratio.

#### 4.2.3 Error analysis

The Kalman filter keeps track of the error estimates of each grid cell emission. The error of the emission analysis depends on the sample frequency of the downwind concentration

	Title:	Algorithm Theoretical Baseline
	Issue:	3.4
GLOBEMISSION	Date: Page:	31 March 2016 32 of 64

plume, the average sensitivity of the emission to the observed concentration, and the error growth of the emissions between assimilation moments. In Figure 4 the emission analysis errors of the entire domain are plotted against the emissions, and are modelled according to equation (29). For OMI data assimilation, we find a dominating absolute error of 0.38  $10^{15}$  molecules/cm<sup>2</sup> at low emissions, and a dominating relative error of 47% at high emissions. Data assimilation with GOME-2 results in larger emission errors due to the smaller amount of observations: 0.48  $10^{15}$  molecules/cm<sup>2</sup> at low emissions, and 74% at high emissions. Note that when the emissions converge to their true state the analysis errors become overestimated due to an overestimation of the emission model covariance **Q**. However, inflating the emission forecast covariance by a significant **Q** at every assimilation cycle is necessary to avoid becoming insensitive to emission changes.

Calculating the emission for an area larger than a grid cell reduces the associated error considerably. For example, for the individual Beijing grid cells in the OMI time series of May-December 2008 the mean relative emission error is 58%. By taking an average over 4 grid cells, the mean error drops to 23%. By taking the negative covariances between the grid cells into account the emission analysis error drops further to 20%. For the GOME-2 time series, the errors are 91%, 47%, and 38%, respectively. The effect of the negative covariances is here stronger for GOME-2 due to the larger footprint size..



Figure 2: S catter plot for emission analysis against its error, for all emissions in the domain at 1 December, 2008. Left panel shows results for OMI, right panel for GOME-2.

#### 4.2.4 DECSO algorithm and emission data versions

#### **DECSO** version 1

- CHIMERE V2006
- No biogenic emissions
- INTEX-B emission inventory
- Landuse by GLCF database (1993)
- European diurnal cycle
- Boundary conditions LMDzINCA (gas), GOCART (aerosol)



#### Used in:

- Mijling and Van der A (2012)
- GlobEmission, phase 1 (East Asia)

#### **DECSO version 2**

- Modelling CHIMERE with sector-dependent emission injection heights
- Fast back-trajectory calculations
- Adjusted retrieval error tropospheric NO2
- Update of NOx-correlated pollutants
- Noise and bias reduction in remote areas
- Full Kalman matrix inversion using LDL decomposition
- Initial emission inventories:
  - o South Africa: EDGAR v4.2, 2008
  - o India: EDGAR v4.2, 2008

#### Used in:

- Mijling et al. (2013)
- GlobEmission, phase 1 (India, South Africa)

#### **DECSO** version v3a

- CHIMERE V2006.
- Initial emission inventories:
  - East Asia: MEIC 2008 (China) + INTEX-B (outside China)
  - South Africa: EDGAR v4.2, 2008
  - o Middle East: EDGAR v4.2, 2008
- Diurnal cycle: flattened for China, European for other regions.
- Calculation speed: switching from g95 to ifort compiler, and calculating matrix inversions with LAPACK

#### Used in:

• GlobEmission, phase 2 (East Asia, Middle East, South Africa)

The emission estimates show unrealistic day-to-day (and possibly month-to-month) fluctuations of emissions. Emission noise in low-emitting areas (introduction of positive emission bias, and unrealistic seasonal cycle when assimilating OMI measurements).

#### **DECSO** version v3b

- CHIMERE V2013: new transport schemes, secondary organic aerosol chemistry, updated chemical reaction rates.
- New land use data: GlobCover Land cover (2009).
- Biogenic emissions by MEGAN
- Initial emission inventories:



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:34 of 64

- East Asia: MEIC 2010 (China) + INTEX-B (outside China), regridded.
  Middle East: HTAP v2 (EDGAR v4.3)
- Reduction of day-to-day emission fluctuations by OmF criterium [-5,10].
- Diurnal cycle: flattened for Middle East, European for East Asia.

#### Used in:

- Ding et al. (2015)
- GlobEmission, phase 2 (East Asia, Middle East)

#### **DECSO version 4**

- Reduction of day-to-day emission fluctuations by 3-sigma (emission error) criterion
- New parametrization of R matrix
- Diurnal cycle: flattened

The emission noise is greatly reduced. The regional emission totals go down, but this is mainly related to the emission noise reduction (reduction of positive bias). Individual hot-spots, especially when undersampled, can disappear (e.g. Ulaanbaatar, and some power plants in North-East China). Apparently a slower convergence rate at changing emission signals.

#### Used in:

• GlobEmission, phase 2 (East Asia)

# 4.3 Emission Inventories for European products

Establishing (long term) trends in pollutant emissions and concentrations is a key part of evaluating the impact of policies. Traditionally, concentrations of air pollutants are monitored using in-situ measurement networks (Tørseth et al., 2012), whereas emissions are estimated on annual basis within the convention for long range transport and air pollution (CLRTAP). Establishing trends based on monitoring networks is hampered by different equipment used at individual sites or countries, replacement of instruments, etc. (Cooper, Gao, Tarasick, Leblanc, & Sweeney, 2012; Sicard, Coddeville, & Galloo, 2009). Moreover, large areas in (south eastern) Europe are not covered by these networks. Similarly, the approaches and quality of emission reporting are variable among European countries (Pouliot et al., 2012). Hence, trend analyses based on satellite datamay provide a valuable independent source of information to compliment traditional monitoring strategies (Bovensmann et al., 1999; Levelt et al., 2006).

To determine the trends in anthropogenic NOx across Europe two approaches are investigated. In the first approach, a fit model is applied to the timeseries of the bias between modelled and retrieved NO2 tropospheric column. It is assumed that the NO2 column change is representative for an emission change and its validity assumption has been tested by Schaap et al. (2013). In the second approach Ensemble Kalman Filter is applied to ingest the retrieved NO2 tropospheric column in the LOTOS-EUROS model. This approach allow for a better capture of the variability in emission strengths as they allow for the estimation of emission variability that changes sign or slope within the period of interest.



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:35 of 64

# 4.3.1 Input data

# 4.3.1.1 Meteorological data

The model is driven by 3-hourly meteorological data provided by European Centre for Medium-range Weather Forecast (ECMWF). These include 3D fields for wind direction, wind speed, temperature, humidity and density, substantiated by 2-d gridded fields of mixing layer height, precipitation rates, cloud cover and several boundary layer and surface variables. The vertical velocity field is calculated using the horizontal wind fields and the mass conservation law of incompressible fluids. Further, the water vapour concentration is calculated using the Claussius-Clapeyron relation. Rain is neglected when the 3-hour accumulated amount is less than 0.3 mm. Linear interpolation is used to derive the meteorological fields at the interval times between the update times.

# 4.3.1.2 Emissions

The major driver of the LOTOS-EUROS system is the anthropogenic emission data of VOC,  $SO_x$ ,  $NO_x$ ,  $NH_3$ , CO,  $CH_4$  and PM. The anthropogenic emission were prescribed following the MACC emission database ([Kuenen et al., 2011). The annual emission totals reported for 2005 have been converted to hourly emission estimates using time factors for the emissions strength variation over the months, days of the week and the hours of the day and scaled for all year between 2003-2007. To account for the occasional fire events, the MACC global fire assimilation system, [Kaiser et al., 2009], is used on a hourly basis. The biogenic NMVOC emissions are calculated online following (Steinbrecher et al., 2009) and the sea salt emissions are parameterised following source formulations for coarse (Monahan et al., 1986) and fine (Mårtensson et al., 2003) aerosol modes.

# **4.3.1.3 Boundary conditions**

The model is constrained by boundary conditions used from the global MACC reanalysis (Hollingsworth et al., 2008, Schere et al., 2012) based on a coupling of the ECMWF model [Flemming et al., 2009] to the chemical transport model MOZART [Emmons et al., 2010].

# 4.3.2 Algorithm description

# **4.3.2.1 LOTOS-EUROS**

The LOTOS-EUROS model is an operational air quality/chemical transport model of intermediate complexity focused on modelling the lower part of the troposphere. Below a description is given of the model characteristics. In this study LOTOS-EUROS model (version 1.10.001), the domain spans from 35° to 70°Nord and -10°to 60°East with a grid resolution of 0.5°longitude x0.25°latitude (approximately 25x25km at mid latitude).

In the vertical the model has four layers up to 3.5 km above sea level: a fixed surface layer of 25 meter and three dynamic layers. The lowest dynamic layer is the mixing layer, followed by two equally thick reservoir layers up to the model top. The height of the mixing layer is part of the meteorological input data. The height of the reservoir layers has a minimum of 50m. In some cases when the mixing layer extends near or above 3.5 km the top of the model

Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	36 of 64

exceeds the 3.5 km according to the abovementioned description. For output purposes the concentrations at measuring height (reference height is usually 3.6 m) are diagnosed by the constant flux approach which relates the dry deposition speed and the concentration of a pollutant.

The chemistry is parameterized following the TNO CBM-IV scheme [Schaap et al., 2008] and the aerosol chemistry is accounted for using the ISORROPIA parameterization (Fountoukis & Nenes, 2007). The transport is represented by advection in three dimension, vertical diffusion and entrainment. The dry deposition is parameterized following following the resistance approach (Erisman et al., 1994, Wichink Kruit et al., 2012). The wet deposition process is represented by below cloud scavenging for gases (Schaap et al., 2004) and particles (Simpson et al., 2003)

# **4.3.2.2** Trends analysis: Remnant approach

For the trend analysis the satellite (OMI and GOME-2) observations are used in synergy with the LOTOS-EUROS model. A multi-year simulation is performed using a single and thus constant a-priory anthropogenic  $NO_x$  emission database. We use the model to estimate the variability due to synoptic variability in weather systems, which it resolves quite well. Moreover, the model resolves part of the seasonal variability due to changing in chemistry and mixing. As we are using a fixed  $NO_x$  emission database the bias between the observations and model results will systematically change as the real emission strength is changing. The changes in the bias can therefore provide fruitful information concerning changes in  $NO_2$  tropospheric columns arising from changes in the emission in the recent years.

In this study we apply a fit model (Weatherhead et al., 1998) to the time series of the bias between model and observation (hereafter called the remnant). In previous studies this fit model was used on  $NO_2$  tropospheric columns using GOME and SCIAMACHY data at a global scale and over China (van der A et al., 2006, van der A. et al., 2008). The fit model can be described by the following function:

$$Y_t = C + \frac{1}{12}Bt + A\sin\left(\frac{\pi}{6}t + \alpha\right) + N_t$$

where  $Y_t$  is the remnant at month t. The first two terms represent a linear trend with slope B representing the annual change in NO<sub>2</sub>. The third term describes the seasonal component of the annual cycle in de bias with amplitude A and phase shift  $\alpha$ .

 $N_t$  is the remainder which cannot be explained by the fit model. The autocorrelation,  $\phi$ , in the remainder affects the precision of the trend. The autocorrelation in the remainders was analysed and an average autocorrelation of 0.1 was found without an indication of a spatial pattern. Hence, the remainders are weakly correlated and the trend precision is determined using the average value. The linear trend B is considered as statistically significant with a 95 % confidence level when  $|B/\sigma_B| > 2$ , where  $\sigma_B$  is the precision of the trend.  $\sigma_B$  is defined as a function of the autocorrelation, the length of the dataset in months and the variance of the remainder,  $\sigma_N$ :



$$\sigma_B = \frac{\sigma_N}{n^{3/2}} \sqrt{\frac{1+\varphi}{1-\varphi}}$$

The fitting model (Weatherhead et al., 1998) is applied to each grid cell across the domain of interest, providing a spatial distribution of the fitting parameters.

Figure 5 shows an example of a measured timeseries and the fitted function. The monthly averaged  $NO_2$  tropospheric column is plotted as a function of the month number starting in January 2005. The black dots represents the measurement while the red and blue lines represent respectively the linear decrease and seasonal component of the fitting result.



Figure 3: An example of a timeseries for one grid cell near Paris (France). The Y axis shows the monthly mean NO2 tropospheric column, and the X axis shows the month index starting January 2005. The black dots represent the measurements from OMI. The red and blue line represent respectively the linear decrease and the seasonal component of the fitting result.

#### **4.3.2.3 Trends analysis: Kalman filter approach**

The trends analysis of the bias assumes that the NO2 column changes is representative for an emission change. For most of the larger countries the assumption is valid, however for smaller countries and in the case of a strong increase in specific sectors such as international shipping the trends in the columns may underestimate actual emission trends of a country. To account for these issues, the chemistry transport model can be used to a larger. In practice, LOTOS–EUROS is equipped with a data assimilation package (Barbu, Segers, Schaap, Heemink, & Builtjes, 2009; Curier et al., 2012), which is used to assimilate the OMI NO2 tropospheric column. To estimate the change in NO<sub>x</sub> emission from year to year, data-assimilation of OMI tropospheric NO<sub>2</sub> measurements in the LOTOS-EUROS chemistry transport model will be performed. The variability in the NOx emission scaling factor from the EnKF should represent the inter-annual changes in the NOx emissions.



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:38 of 64

#### Assimilation system

A data assimilation system has been developed around the LOTOS-EUROS model based on the Kalman filter technique. A Kalman filter computes probability density functions (pdfs) of the true state, given 1) a transition model to propagate the state in time with associated uncertainties; and 2) observations with associated representation error. Starting from initial pdf, the filter first performs a *forecast* step propagating the pdf in time to the first moment that observations become available. Then, during the *analysis* step, the forecasted pdf is replaced by an analyzed version that takes into account the new information that has become available. The Kalman filter is an example of a sequential assimilation, since forecast and analysis steps follow each other sequentially in time and use only information from the past.

The original linear Kalman Filter is closed, i.e. if the initial pdf, the model uncertainty, and the representation error are expressed as Gaussian (normal) random vectors, and if in addition the transition model is linear, then the computed pdf's for the true state are Gaussian too. The advantage of Gaussian random vectors is that the complete pdf is described by a mean state vector and a covariance matrix only. In practice, the linear Kalman filter cannot be implemented for large scale applications as the size of the state vector is usually very large (at least  $n=10^4$  elements) and storage of a covariance matrix becomes impossible since it requires  $n^2$  elements. In addition, the linear Kalman filter requires 2n evaluations of the transition model, which is far too expensive for the type of models considered here. Moreover transition models are usually non-linear as chemical reactions are included.

A suitable alternative for the linear Kalman filter for air pollution models is the Ensemble Kalman filter (EnKF)(Evensen 1994). In this formulation, the pdf of the state is not expressed in terms of a mean and covariance only, but is described by an ensemble of model states. The spread between the ensemble members should describe the uncertainty in the value of the state and quantities as mean and covariance of the state are computed from the ensemble statistics. The transition model is therefore not restricted to linearity, which simplifies the actual formulation and allows for a transparent implementation. The number of required ensemble members depends on the complexity of the pdf to be described, which is usually determined by the non-linearity of the transition model and the complexity of the associated model uncertainty. In practice, an ensemble with 10-100 is acceptable to keep computations feasible, and the complexity of the problems is limited to a point that this number is reached.

In the EnKF formulation around the LOTOS-EUROS model used in this study, the basic time step between two analyses is one hour. At the end of each step, all observations that have become available are analyzed. The number of analyzed observations therefore depends on the overpass time of the satellites.

#### Transition model

The transition model from time (day) t[k] to t[k+1] is given by:



$$\begin{pmatrix} c[k+1]\\\lambda[k+1] \end{pmatrix} = \begin{pmatrix} M(c[k],t[k],\lambda[k])\\\alpha\lambda[k] \end{pmatrix} + \begin{pmatrix} o\\\sqrt{1-\alpha^2}\sigma \end{pmatrix}$$
 Eqs  
$$x[k+1] = M(x[k]) + \Gamma w[k]$$

Equation 1

The transition model consists of 2 parts:

- The vector c[k] contains the instantaneous model state at the end of the time step (here hourly). The model state consists of the concentrations of all components in each grid cell, as well as concentration depended fields such as the total aerosol water content. Operator M denotes the propagation of the concentrations by the LOTOS-EUROS model from hour to hour.
- The model is driven by uncertainty parameters  $\lambda[k]$ , which are kept constant during the time step.

In the current application, uncertainty is defined for the emission input. Emissions from anthropogenic as well as natural emissions are considered uncertain, to enforce variation in the studies key species O3 and NO2 but also in HCHO are CO, where the later are strongly influenced by natural emissions. Every time the model simulates emissions, the actual emission of group p in grid cell ij is described by:

$$s_{ij}^{p}[k] = \overline{s}_{ij}^{p}(1 + \lambda_{ij}^{p}[k]),$$

#### Equation 2

where  $\overline{s}$  denotes the nominal value. To account for emissions just outside the domain, the boundary conditions are changed similar to the adjacent emissions. The uncertainty  $\lambda$  is described as a colored noise process driven by an *r*-element uncorrelated white noise w[k] with zero mean and identity covariance; each element accounts for a different emission field. The standard deviation of the uncertainty is set to a value  $\sigma$  which has to be determined. To ensure that only realistic emissions are used, the values of  $\lambda$  are bounded into the interval [-1,1], such that emissions are never non-negative or more than doubled. A temporal correlation is assumed with exponential decay; by this, the value of the emissions will change gradually from day to day, to ensure that emissions are lower or higher than the nominal value during a longer period. The driving white noise w[k] has only *r* elements (one for each group of uncertainty. However, since the noise factors  $\lambda$  are available for each grid cell, spatial variations will be introduced during the analysis step.

The uncertainty specification has a number of undefined settings yet, which have to be defined prior to the assimilation experiments. These settings include the selection of the uncertain emissions (source categories, emitted species), the amplitude of the assumed uncertainties, and their temporal correlation. Appropriate values are those that can explain the observed difference between the model simulations and the observations, taking into account the representation error. These values will therefore be chosen once the (synthetic) observations are available.

#### Ensemble formulation

An ensemble of m state vectors forms the main data structure in the EnKF:



$$\xi_i[k]$$
 ,  $j = 1,..,m$  Equation 3

The spread in the ensemble should at any time represent the uncertainty about the value of the true state x[k]; the true value exists but is unknown to the user, who could only provide an estimate of what are likely values. A mean and covariance for the unknown true state can be computed from the ensemble:

$$\overline{x} = \frac{1}{m} \sum_{j=1}^{m} \xi_j$$
Equation 4
$$P = \frac{1}{m-1} \sum_{j=1}^{m} (\xi_j - x) (\xi_j - x)^T$$
Equation 5

The ensemble members are initialized at day k=0 with 'analysed' filter states  $\xi_j^a[0]$ , which are all the same and include model states c[0] resulting from a model run over the two weeks prior to the assimilation period, and uncertainty parameters  $\lambda[0]$  drawn from random generator with zero mean and standard deviation  $\sigma$ . In the forecast stage of the filter, each ensemble member  $\xi_j$  is propagated over the next time step using a white noise sample  $w_j$  drawn from a random generator:

$$\xi_{i}^{f}[k] = M(\xi_{i}^{f}[k-1]) + \Gamma w_{i}[k-1]$$

#### Observation representation

At the end of each step, the (synthetic) observations that have become available during this time step are collected in a vector y[k]. The observations can be simulated from a filter state by an operator H; the operator extracts from the concentration array the value at the station location, or the (partial) column over the footprint of a satellite pixel taking into account the averaging kernel. The difference between the observations and the simulations from the (unknown) true state is supposed to be described by:

$$y[k] - H[k]x[k] = v[k]$$
 Equation 6

where v[k] is a random vector with zero mean covariance R[k]. These representation errors (or residues) are supposed to be uncorrelated (*R* is diagonal). The standard deviations that define the diagonal have to be determined by examining the difference between the simulations and the (synthetic) observations, taking into account the model uncertainty to (see previous section on uncertainty model). A parameterization for the representation error standard deviation could be that it is equal to a fraction of the observed value, with thresholds for the minimum and maximum values.

For representation of satellite observations, also the size of the footprints with respect to the model grid cell size should be taken into account. The spatial representation error that is made by averaging concentrations over the footprint could be determined by comparing model simulations at the highest resolution with fields obtained after spatial averaging. This procedure has been applied to the LOTOS-EUROS simulations of formaldehyde of the nature run; the results are shown in Figure 1.





Figure 1:Variance of difference between high resolution simulations of formaldehyde (7 km) and averages over surrounding grid cells, as a function of the horizontal average. Given the size of a satellite footprint, this curve provides the spatial averaging error that arises from averaging over footprints.

#### Analysis

The ensemble is analysed using the available observations. Since the observation errors are supposed to be uncorrelated, the observations can be analysed one-by-one. For observation l, the first step in the analysis is to compute a gain matrix:

$$K_{l} = \widetilde{P}^{f} H^{T} \left( H \widetilde{P}^{f} H^{T} + r_{ll} \right)^{-1}$$
 Equation 7

where  $r_{ll}$  is the diagonal element of R for observation l; the time indices are omitted to simply the notations. Since only one observation is analysed here, the inverse at the right hand side is simply a division by a scalar. The matrix  $\tilde{P}^{f}$  is a covariance matrix computed from the ensemble. Simply using the ensemble covariance for this is undesirable, since the spatial correlations represented by the ensemble are usually too strong. This exaggeration is mainly caused by the use of a finite ensemble size, which is useful to represent the major part of the covariance but is unable to represent all details correctly. In addition, the uncertainty model for the emissions has no spatial variation and therewith introduces an artificial increase of spatial correlations. Thirdly, even though their emission origin and timing may be independent, a similar diurnal cycle in the concentrations can create artificial or spurious correlations. As a consequence, the ensemble covariance may suggest that grid cells which are far apart from each other in the domain are strongly correlated, which is undesirable since these correlations are used to distribute the residue between an observation and a simulation over the domain. To be able to ignore the spurious correlations, the localization procedure described in [Houtekamer and Mitchell, 2001] is used. In this procedure, the covariance matrix actually used is formed from a Schur product between the ensemble covariance and a correlation matrix with finite band width. An element  $\tilde{p}_{ii}$  of this covariance matrix is computed using:

$$\widetilde{p}_{ij} = p_{ij} C(d_{ij})$$
 Equation 8

where  $d_{ij}$  is the horizontal distance between the grid cells holding elements *i* and *j*, and C(d) is a spatial correlation function. An appropriate formulation for C(d) is determined by examining the spatial correlations present in the simulation (Figure 2). At large distances, the



correlation factor vanishes and is explicitly set to zero. As a result, the gain matrix is zero for all elements outside this range, which limits the computational costs since only a small part of the full gain needs to be computed.

With the (localized) gain for observation l available, the ensemble members are analysed by:

$$\xi_{j;l}^{a} = \xi_{j;l-1}^{a} + K_{l} \left( y_{l} - H_{l} \overline{x}_{l-1}^{a} \right)$$
 Equation 9

where  $\xi_{j;l-1}^{a}$  is the ensemble member resulting from the analysis of the previous observation, and similar  $\bar{x}_{l-1}^{a}$  is the mean state computed over all previously analysed members; for the first observation, these are set to the forecast entities. The effect of the analysis is that elements of the state are changed following their relation to simulated observations.



Figure 4 S patial correlation in formaldehyde simulations in the nature run. Each dot represents a correlation between two time series of concentrations from different grid cells, plotted against the spatial distance between the cells. The averages (bars) are used to fit an appropriate spatial correlation function that can be used for the localization procedure.

#### Observation screening

In case the assimilation system is unable to represent an observation correctly, assimilation could lead to instability of the system. Such instability can occur if the model lacks certain physical parameterizations or if the model is unable to represent a measurement as a result of a mismatch of the model and measurement spatial or temporal resolution. To avoid this, a screening procedure is applied to reject those measurements which cannot be represented correctly by the assimilation system. The screening procedure is taken from [Jarvinen and Unden, 1997]. If the square of the difference between observation and filter mean is much larger than the expected variance of this difference, the observation is rejected. That is:

$$\left(y_{l}-H_{l}x^{f}\right)^{2} > \alpha\left(H_{l}P^{f}H_{l}^{T}+r_{ll}\right)$$
 Equation 10

where the parameter  $\alpha$  is a threshold factor. The rejected observations are flagged to be able to investigate afterwards where the filter is unable to represent them.

Ensemble size

Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	43 of 64

An important parameter in the filter that has not been discussed yet is the ensemble size. In general, the ensemble size should be large enough to represent the covariance structure imposed by model uncertainty and the model physics. In this application, the covariance structure is rather simple, since all uncertainty is described by the local sources, and observations are available regularly over the domain. Experiments showed that for the described application a limited number of 12-15 modes is sufficient. The current assimilation system is based on conclusions from an extensive sensitivity study carried out within the framework of the SMOGPROG project [Swart et al., 2008], a number of 12 is currently used for the operational forecasts.

#### 4.3.3 Error analysis

tba



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:44 of 64

# 4.4 Emission Estimates related to Aerosols

#### 4.4.1 Input data

The input data for the aerosol emission estimation will be based on AOD retrievals of AATSR and MODIS.

# 4.4.2 Algorithm description

The emission inversion is based on using the SILAM chemistry transport model driven by 4dimensional variational inversion scheme.

System for Integrated modeLling of Atmospheric coMposition SILAM (Sofiev et al., 2006, 2008) currently includes both Eulerian and Lagrangian advection-diffusion formulations. The Eulerian core used in the current study is based on the transport scheme of Galperin (1999, 2000), which incorporates the horizontal diffusion term and is combined with the extended resistance analogy of Sofiev (2002) for vertical diffusion.

The removal processes are described via dry and wet deposition. Depending on the particle size, mechanisms of dry deposition vary from the primarily turbulent diffusion driven removal of fine aerosols to the primarily gravitational settling of coarse particles (Slinn and Slinn, 1980). The SILAM wet deposition parameterization (Sofiev et al., 2006, Horn et al., 1987, Smith and Clark, 1989, Jylha, 1991) is based on direct observations performed for moderately hydrophobic aerosols. It distinguishes between sub- and in-cloud scavenging by both rain and snow. The particle size dependence of the impaction scavenging is taken into account by increasing the scavenging rate for super-micron particles in relation to their settling velocity.

The system includes a meteorological pre-processor for diagnosing the basic features of the boundary layer and the free troposphere from the meteorological fields provided by numerical weather prediction (NWP) models (Sofiev et al., 2010). Physical-chemical transformation modules of SILAM include several tropospheric chemistry schemes, basic aerosol dynamics, and radioactive decay processes. The system accepts flexible definition of the particle size spectrum, which can be defined for each specific run depending on the application.

The emission inversion is based on a variant of the 4D-Var data assimilation method, which involves iterative minimization of a quadratic cost function depending on the distance of the observations and their modeled counterparts, and on the distance between the emission estimate and the a priori emission inventory. A detailed description of the emission inversion method has been given in Vira and Sofiev (2012). In the following, we present a summary of the method, and its extension into particulate matter.

Let us denote the parameter (such as initial state or emission rate) of interest as  $\xi$ , and define the model operator *M* mapping the parameter, or the control variable, to a unique phase-space trajectory  $x=M\xi$  defined over some finite time interval referred as the assimilation window.

Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	45 of 64

The vector of observations y corresponds to the model state x via the observation operator H:  $y = H(x) + \varepsilon$ , where  $\varepsilon$  is the observation error, which is assumed to be Gaussian.

The maximum likelihood estimate of the parameter  $\xi$  is then the value minimising the cost function

$$J(\xi) = \frac{1}{2} (y - Hx)^T R^{-1} (y - Hx) + \frac{1}{2} (\xi - \xi_b)^T B^{-1} (\xi - \xi_b)$$
(1)

The first term penalises the deviation from the observations *y* whose accuracy is described by the covariance matrix *R*. The prior knowledge of  $\xi$  is included in the background value  $\xi_b$  and the background error covariance matrix *B*. The second term therefore penalises the deviation from the prior  $\xi_b$ . The cost function is minimised using iterative numerical algorithms.

The gradient of equation (1) with respect to  $\xi$  is

$$J'(\xi) = M^* H^* R^{-1} (y - Hx) + B^{-1} (\xi - \xi_b),$$
(2)

where  $M^*$  and  $H^*$  are the tangent linear adjoint model and observation operators, respectively. (Marchuk, 1995).

The forward dispersion model corresponding to the operator M and defining the time evolution of the model state is defined by the scalar transport equation

$$\frac{\partial c_n}{\partial t} + \frac{\partial}{\partial x_i} (u_i c_n) = \frac{\partial}{\partial x_i} \mu_{ii} \frac{\partial c_n}{\partial x_i} + S_n(x,t) + f_n(x,t)$$
(3)

where  $c_n$  is a concentration of the *n*-th species,  $f_n(x,t)$  is the emission density, and the chemical sources and sinks are included in S(c,t). If the reaction term is linear, i.e.  $S(c,t) = kc_n(x,t)$ , then the adjoint equation to (3) reads as (Marchuk, 1995)

$$\frac{\partial c_n^*}{\partial t} - \frac{\partial}{\partial x_i} (u_i c_n^*) = \frac{\partial}{\partial x_i} \mu_{ij} \frac{\partial c_n^*}{\partial x_j} + k c_n^* (x, t)$$

Here  $c^*(x,t)$  is the first-order sensitivity of the functional (1) to a concentration perturbation at time t. Its solution corresponds to  $M^*$  in (2).

Estimating the complete four dimensional emission distribution is impractical due to underdetermination of the problem. Therefore, the approach used in this work and shared by previous studies (Yumimoto *et al.*, 2007, Elbern *et al.*, 2007) is to assume a constant relative deviation of the emission intensity from the background rate throughout the assimilation and forecast windows. The adjusted emission rate is written as  $f(x,t) = \alpha(x)f_0(x,t)$ , where  $\alpha(x)$  is to be estimated. The diurnal emission variations are thus not affected by the assimilation. The correction factor  $\alpha(x)$  is assumed to be constant along the vertical axis (height). The sensitivity, and, consequently, the gradient of  $J(\xi)$  with respect to  $\alpha(x)$  (see equation (2)) is obtained by integrating the solution of the adjoint problem (4) over the assimilation window and the vertical extent of the model domain.

The final emission estimates are given by multiplying the prior emission field with the factors  $\alpha$ . An assimilation window of 24 hours will be used for the GlobEmission aerosol product.



While the system therefore produces the estimates on daily level, better robustness is expected from the monthly averaged values.

The primary source of observational information for the aerosol emission estimates are satellite retrievals of the aerosol optical depth (AOD). The AOD observation operator is defined by the profile of the mass extinction coefficient, which is computed using the Mie theory for a prescribed particle size distribution. The particle growth due to humidity is considered following Sofiev *et al*, (2011).

The SILAM model currently includes the emission and formation of primary organic and inorganic, as well as secondary inorganic aerosols. However, the computational cost of the emission inversion requires a simplified representation of the heterogeneous chemistry. The emission inversion is therefore carried out in two steps: first, the contribution of non-sulphuric aerosol compounds (NH4<sup>+</sup>, NO<sub>3</sub><sup>-</sup>) to the total AOD is computed, and subtracted from the observed values. Then, the optimization problem (1) is solved using a reduced model including only natural and anthropogenic primary aerosols and sulphate formation.

# 4.4.3 Error analysis

#### • AATSR

The uncertainty estimation for the resulting AOD is obtained by propagating the measurement error of the top-of-atmosphere reflectance by applying inverse problem theory. After the result is determined, the uncertainty is computed using a posteriori covariance. The measurement error is taken to be 5 % of the measured reflectance and it is assumed to be uncorrelated. (ESA- CCI project)

#### • MODIS

A metric analysis of the retrieved AOD values is the tolerance which bounds AOD errors. Hyer et al., 2011 define the target accuracy as:

$$\tau_{_{M}} = \tau_{_{\rm A}} \pm (0.05 + \frac{\tau_{_{\rm A}}}{5}) \tau_{_{\rm coarse}}$$

Globally, one standard deviation of "very good" data falls within  $(0.05 + 0.2 \times \tau_A)$  error thresholds ( $\tau_A$  is the AOD at the MODIS retrieval at 555 nm). However, for most applications a prognostic RMS (root mean square) error model with a noise floor is more appropriate. For global applications using MODIS level 2 data over land, we recommend the use of the greater of 0.08 or  $0.02 + 0.22 \times \tau_M$  for terra and 0.07 or  $0.01 + 0.26 \times \tau_M$  for Aqua. (Shi et al., 2011, Zhang et al 2006, Remer et al 2005, among others). Where  $\tau_M$  is "Corrected Optical Depth-Land" from the level 2 product.

The emission estimates are affected by the uncertainties in observations, a priori emissions as well as uncertainties in the chemistry-transport model. The observational and a priori emission errors are encoded in the covariance matrices R and B in Eq. (1). While an algebraic relation between R, B, and the a posteriori errors is easy to derive, its evaluation is computationally difficult due to the dimension of the problem. Moreover, the effect of forward model errors is not quantified this way.

Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	47 of 64

The sensitivity to errors in the input data will be studied using Monte-Carlo simulations with synthetic observations. The effect of modeling errors can be investigated with cross-validation using several chemistry-transport models with independent inversion schemes.



# 4.5 Emission Estimates related to fires

# 4.5.1 Input data

The input data for the aerosol emission estimation will be based on AOD retrievals of AATSR and MODIS.

# 4.5.2 Algorithm description

The emission inversion is based on using the SILAM chemistry transport model driven by 4dimensional variational inversion scheme.

System for Integrated modeLling of Atmospheric coMposition SILAM (Sofiev et al., 2006, 2008) currently includes both Eulerian and Lagrangian advection-diffusion formulations. The Eulerian core used in the current study is based on the transport scheme of Galperin (1999, 2000), which incorporates the horizontal diffusion term and is combined with the extended resistance analogy of Sofiev (2002) for vertical diffusion.

The removal processes are described via dry and wet deposition. Depending on the particle size, mechanisms of dry deposition vary from the primarily turbulent diffusion driven removal of fine aerosols to the primarily gravitational settling of coarse particles (Slinn and Slinn, 1980). The SILAM wet deposition parameterization (Sofiev et al., 2006, Horn et al., 1987, Smith and Clark, 1989, Jylha, 1991) is based on direct observations performed for moderately hydrophobic aerosols. It distinguishes between sub- and in-cloud scavenging by both rain and snow. The particle size dependence of the impaction scavenging is taken into account by increasing the scavenging rate for super-micron particles in relation to their settling velocity.

The system includes a meteorological pre-processor for diagnosing the basic features of the boundary layer and the free troposphere from the meteorological fields provided by numerical weather prediction (NWP) models (Sofiev et al., 2010). Physical-chemical transformation modules of SILAM include several tropospheric chemistry schemes, basic aerosol dynamics, and radioactive decay processes. The system accepts flexible definition of the particle size spectrum, which can be defined for each specific run depending on the application.

The emission inversion is based on a variant of the 4D-Var data assimilation method, which involves iterative minimization of a quadratic cost function depending on the distance of the observations and their modeled counterparts, and on the distance between the emission estimate and the a priori emission inventory. A detailed description of the emission inversion method has been given in Vira and Sofiev (2012). In the following, we present a summary of the method, and its extension into particulate matter.

Let us denote the parameter (such as initial state or emission rate) of interest as  $\xi$ , and define the model operator *M* mapping the parameter, or the control variable, to a unique phase-space trajectory  $x=M\xi$  defined over some finite time interval referred as the assimilation window.

Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	49 of 64

The vector of observations y corresponds to the model state x via the observation operator H:  $y = H(x) + \varepsilon$ , where  $\varepsilon$  is the observation error, which is assumed to be Gaussian.

The maximum likelihood estimate of the parameter  $\xi$  is then the value minimising the cost function

$$J(\xi) = \frac{1}{2} (y - Hx)^T R^{-1} (y - Hx) + \frac{1}{2} (\xi - \xi_b)^T B^{-1} (\xi - \xi_b)$$
(1)

The first term penalises the deviation from the observations *y* whose accuracy is described by the covariance matrix *R*. The prior knowledge of  $\xi$  is included in the background value  $\xi_b$  and the background error covariance matrix *B*. The second term therefore penalises the deviation from the prior  $\xi_b$ . The cost function is minimised using iterative numerical algorithms.

The gradient of equation (1) with respect to  $\xi$  is

$$J'(\xi) = M^{+}H^{+}R^{-1}(y - Hx) + B^{-1}(\xi - \xi_{b}), \qquad (2)$$

where  $M^*$  and  $H^*$  are the tangent linear adjoint model and observation operators, respectively. (Marchuk, 1995).

The forward dispersion model corresponding to the operator M and defining the time evolution of the model state is defined by the scalar transport equation

$$\frac{\partial c_n}{\partial t} + \frac{\partial}{\partial x_i} (u_i c_n) = \frac{\partial}{\partial x_i} \mu_{ii} \frac{\partial c_n}{\partial x_i} + S_n(x,t) + f_n(x,t)$$
(3)

where  $c_n$  is a concentration of the *n*-th species,  $f_n(x,t)$  is the emission density, and the chemical sources and sinks are included in S(c,t). If the reaction term is linear, i.e.  $S(c,t) = kc_n(x,t)$ , then the adjoint equation to (3) reads as (Marchuk, 1995)

$$\frac{\partial c_n^*}{\partial t} - \frac{\partial}{\partial x_i} (u_i c_n^*) = \frac{\partial}{\partial x_i} \mu_{ij} \frac{\partial c_n^*}{\partial x_j} + k c_n^* (x, t)$$

Here  $c^*(x,t)$  is the first-order sensitivity of the functional (1) to a concentration perturbation at time t. Its solution corresponds to  $M^*$  in (2).

Estimating the complete four dimensional emission distribution is impractical due to underdetermination of the problem. Therefore, the approach used in this work and shared by previous studies (Yumimoto *et al.*, 2007, Elbern *et al.*, 2007) is to assume a constant relative deviation of the emission intensity from the background rate throughout the assimilation and forecast windows. The adjusted emission rate is written as  $f(x,t) = \alpha(x)f_0(x,t)$ , where  $\alpha(x)$  is to be estimated. The diurnal emission variations are thus not affected by the assimilation. The correction factor  $\alpha(x)$  is assumed to be constant along the vertical axis (height). The sensitivity, and, consequently, the gradient of  $J(\xi)$  with respect to  $\alpha(x)$  (see equation (2)) is obtained by integrating the solution of the adjoint problem (4) over the assimilation window and the vertical extent of the model domain.

The final emission estimates are given by multiplying the prior emission field with the factors  $\alpha$ . An assimilation window of 24 hours will be used for the GlobEmission aerosol product.



While the system therefore produces the estimates on daily level, better robustness is expected from the monthly averaged values.

The primary source of observational information for the aerosol emission estimates are satellite retrievals of the aerosol optical depth (AOD). The AOD observation operator is defined by the profile of the mass extinction coefficient, which is computed using the Mie theory for a prescribed particle size distribution. The particle growth due to humidity is considered following Sofiev *et al*, (2011).

The SILAM model currently includes the emission and formation of primary organic and inorganic, as well as secondary inorganic aerosols. However, the computational cost of the emission inversion requires a simplified representation of the heterogeneous chemistry. The emission inversion is therefore carried out in two steps: first, the contribution of non-sulphuric aerosol compounds (NH4<sup>+</sup>, NO<sub>3</sub><sup>-</sup>) to the total AOD is computed, and subtracted from the observed values. Then, the optimization problem (1) is solved using a reduced model including only natural and anthropogenic primary aerosols and sulphate formation.

# 4.5.3 Error analysis

#### • AATSR

The uncertainty estimation for the resulting AOD is obtained by propagating the measurement error of the top-of-atmosphere reflectance by applying inverse problem theory. After the result is determined, the uncertainty is computed using a posteriori covariance. The measurement error is taken to be 5 % of the measured reflectance and it is assumed to be uncorrelated. (ESA- CCI project)

#### • MODIS

A metric analysis of the retrieved AOD values is the tolerance which bounds AOD errors. Hyer et al., 2011 define the target accuracy as:

$$\tau_{\scriptscriptstyle M} = \tau_{\scriptscriptstyle \rm A} \pm (0.05 + \frac{\tau_{\scriptscriptstyle \rm A}}{5}) \tau_{\scriptscriptstyle \rm coarse}$$

Globally, one standard deviation of "very good" data falls within  $(0.05 + 0.2 \times \tau_A)$  error thresholds ( $\tau_A$  is the AOD at the MODIS retrieval at 555 nm). However, for most applications a prognostic RMS (root mean square) error model with a noise floor is more appropriate. For global applications using MODIS level 2 data over land, we recommend the use of the greater of 0.08 or  $0.02 + 0.22 \times \tau_M$  for terra and 0.07 or  $0.01 + 0.26 \times \tau_M$  for Aqua. (Shi et al., 2011, Zhang et al 2006, Remer et al 2005, among others). Where  $\tau_M$  is "Corrected Optical Depth-Land" from the level 2 product.

The emission estimates are affected by the uncertainties in observations, a priori emissions as well as uncertainties in the chemistry-transport model. The observational and a priori emission errors are encoded in the covariance matrices R and B in Eq. (1). While an algebraic relation between R, B, and the a posteriori errors is easy to derive, its evaluation is computationally difficult due to the dimension of the problem. Moreover, the effect of forward model errors is not quantified this way.



The sensitivity to errors in the input data will be studied using Monte-Carlo simulations with synthetic observations. The effect of modeling errors can be investigated with cross-validation using several chemistry-transport models with independent inversion schemes.



# 4.6 Emission Estimates related to fires

# 4.6.1 Input data

The Integrated System for wild-land FIRES (IS4FIRES) v.1.2 is based on Level 2 MODIS Collection 5.1 Active Fire Products, which are used for the near-real-time and historical evaluation of the emissions from wild-land fires, as well as a series of other input data (Figure 5). Considerations are given to include fire products of VIIRS, GOES-2 and GOES-3 when these data become available.



Figure 5. Scheme of data flows of IS4FIRES v.1.2.

The incorporation of the fire injection profile resulted in a strict separation of the satelliterelated data containing the fire information and land cover from the meteorology-related data used for calculation of the initial 3D plume distribution. For the sake of efficiency, the second part has been incorporated inside the SILAM model – because this is the first system that naturally involves the meteorological–data processing routines. Therefore, the IS4FIRES v.1.2 per-se became smaller: it only checks, reprojects and aggregates the MODIS fire observations in the daily fire observations. The rest of emission calculation takes place in SILAM, which calculates the absolute emission and its 3D injection clouds dynamically while evaluating the impact of fires on atmospheric composition and air quality.

 Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	53 of 64

The present IS4FIREScconsists of two parallel branches based on partly independent products: the Temperature Anomaly and Fire Radiative Power. Their algorithms of converting the fire information to the emission fluxes of atmospheric pollutants are described below, starting from the outlines of the corresponding fire products. They are used so that, if the FRP information on the specific fire is available, it is taken as-is. Should the FRP processing of MODIS fail but the TA processing pass, the TA value is used for producing the would-be-FRP value following the non-linear regression suggested by Sofiev et al, (2009).

The MODIS fire detection procedure is based on a contextual algorithm of Giglio et al (2003) that exploits the strong emission of mid-infrared radiation from fires (Dozier, 1981; Matson and Dozier, 1981). The algorithm examines each pixel of the MODIS swath and attributes it to one of the following classes: missing data, cloud, water, non-fire, fire, or unknown. For each fire-classified pixel, the procedure attempts to use the neighbouring pixels to estimate the radiometric signal of the pixel, if there would be no fire there. Valid neighbouring pixels are identified in a window centred on the potential fire pixel and used to estimate this background value.

# 4.6.2 Algorithm description

The procedure of determining the emission from the wild-land fires will generally follow that of aerosol products – with a few exceptions. Firstly, the background emission will be taken from IS4FIRES system of FMI outlined below. Secondly, the data assimilation will be used to constrain both particulate matter and gaseous species, with primary attention paid to CO.

The Fire Radiative Power (FRP, a rate of release of Radiation Energy, FRE) of the fire pixel is based on the empirical formula of Kaufman *et al* (1998):

$$FRP = 4.34 * 10^{-13} \left( T_4^8 - T_{4b}^8 \right), \quad [Watt], \tag{1}$$

where the  $T_4$ , and  $T_{4b}$  are the fire and the background (taken from neighbouring pixels) temperatures, respectively, measured at the 4-µm channel. The dependence has been obtained from fitting the actual release of radiative energy from a fire and its apparent temperature at the 4 and 11 µm channels – as observed by the MODIS instrument. The relationship showed good correlation for open moderate-to-strong fires (Kaufman *et al*, 1998). There may be potential difficulties for small fires, as these may be partly overshadowed by trees, appear as low-temperature but strongly emitting smouldering fires, etc.

Both TA and FRP data are included into the level 2 Fire Products (MOD14 for Terra and MYD14 for Aqua satellites).

To convert the FRP to emission fluxes we used a similar approach as for TA – a direct conversion of FRP using an empirical scaling to emission rates. In the current FAS it is based on Ichoku & Kaufman (2005, hereinafter referred as IK05) who related the FRP in [W] per pixel to total particulate matter (PM) emission in [kg tPM s<sup>-1</sup>]. Since the calibration IK05 was obtained by relating the aerosol optical depth (AOD) with the FRP, the obtained emission factors are valid for total PM instead of PM<sub>2.5</sub>, which was the reference species for FAS-TA. The mean relation between these PM measures can be evaluated based on AM01: within the fire plume

 Title: Ref.:	Algorithm Theoretical Baseline GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	54 of 64

 $m_{pm2.5} \approx 0.6 m_{total pm}$ 

(2)

The relation is approximately valid for all land use types: the changes between the vegetation types are smaller than the uncertainty range within each type (AM01).

The key parameter for IS4FIRES is therefore the emission rate of total PM per unit FRP, i.e. the smoke emission factor  $C_e$  [kg tPM J<sup>-1</sup>]. According to IK05,  $C_e$  varies from 0.02-0.06 kg tPM MJ<sup>-1</sup> for boreal regions, 0.04-0.08 kg tPM MJ<sup>-1</sup> for Africa (mainly savannas and grassland), and 0.08-0.1 kg tPM MJ<sup>-1</sup> for Western Russian regions. Since the  $C_e$  determination involved a simple estimate of atmospheric transport (based on wind at a constant height and not involving a dispersion model), the authors suggested that the coefficients are probably overestimated by about a factor of 2. Using these estimates as a starting point, we have re-estimated the emission coefficients using actual land-cover information, instead of geographical regions. For the IS4FIRES v.0.99 and v.1.0, three land cover classes were considered: forest, grass, and mixture of both. The corresponding total-PM emission factors were suggested as: 0.035 kg tPM MJ<sup>-1</sup> for forest, 0.018 kg tPM MJ<sup>-1</sup> for grassland and agriculture, and 0.026 kg tPM MJ<sup>-1</sup> for mixed areas. These values were deduced from the prevailing land cover in the IK05 domains.

For the IS4FIRES v.2, the primary deliverable of the GlobEmission, the list of land cover classes was increased to 7 types (Table 2).

Land cover	PM <sub>2.5</sub> factor	unit	PM <sub>10</sub> factor	Unit
Tropical forest	0.00697	kg/MJ	0.01417	kg/MJ
Grass	0.00410	kg/MJ	0.00474	kg/MJ
Crop residue	0.01003	kg/MJ	0.02083	kg/MJ
Pasture maintenance	0.02058	kg/MJ	0.04020	kg/MJ
Boreal forest	0.01085	kg/MJ	0.01419	kg/MJ
Temperate forest	0.00410	kg/MJ	0.00484	kg/MJ
Peat	0.02086	kg/MJ	0.04173	kg/MJ
Shrub	0.00475	kg/MJ	0.00558	kg/MJ

Table 2. Emission factors for diffe	erent land cover types (intermed	liate calibration based on 2	008 global simulations).

We assume that inside the fire plumes, the AOD was entirely determined by the biomassburning products. Thus, Saarikoski et al (2007) found that more than 80% of  $PM_{2.5}$  during a specific episode in May 2006 was originated from fires – in comparatively aged plumes. We



therefore attributed all systematic discrepancy between the observed and calculated column AOD to errors in the emission rates, and corrected the emission factors accordingly.

The intermediate calibration performed on the basis of a single year 2008 resulted in the emission factors shown in Table 2. For the final calibration, these values will be used as a starting point for refinement via both scaling calibration for other years and variational assimilation over specific episodes and regions, esp0ecially those dominated by a specific land cover type.

# 4.6.3 Error analysis

Direct error analysis of emission estimates is not possible due to no direct observations of this quantity. Therefore, the error analysis will concentrate on two complementary directions: (i) evaluation of the resulting model plumes from fires using the satellite and in-situ pollutant data, (ii) inter-comparison with existing emission inventories for the same time period.

Comparison with existing inventories will be made for evaluating the general uncertainties of the emission estimates rather for correcting the GlobEmission estimates: a-priori it is not known, which inventory is more accurate.

Provision will be given to application of the obtained emission inventory in another model of GlobEmission – LOTOS-EUROS, which would demonstrate the impact of the model formulations to the emission estimations.



# 4.7 High Resolution Emission Inventories

# 4.7.1 Input data

- **Regional emission inventory for Middle East, provided by KNMI:** KNMI will provide emission estimates of NOx on a 0.25 degree resolution, derived from satellite observations and inverse modelling, for the Middle East.
- Sector split provided by QEERI or based on literature search: The regional emission inventory for the Middle East will be provided by KNMI as total emissions, i.e. without any breakdown over different economic sectors. As further downscaling is highly dependent on the breakdown over different sectors, information on the sectoral split is required.

For South Africa (GlobEmission, phase 1), relative values were taken from literature, however these data had a lot of draw-backs: the data were outdated, they were only available at national level, only dealt with anthropogenic emissions, and only 4 main sectors were included, ....

For the downscaling application in Qatar, we hoped to find better data. Hereto, it was agreed with QEERI, the GlobEmission user from Qatar, that VITO would perform a literature search on available data with respect to relative contributions (biogenic versus anthropogenic on one hand and contributions of (sub)sectors within both, biogenic and anthropogenic emissions, on the other hand). Moreover, VITO would perform a screening on the availability of similar data in surrounding areas (Saudi Arabia, United Arab Emirates, ...). QEERI from its side would investigate which data are available via the Qatar national reporting system (Qatar needs to report emissions to the UN). Furthermore, it was also agreed that QEERI would facilitate discussions with eventual data suppliers (e.g. from surrounding areas) on VITOs request.

During the project it became clear that QEERI would be unable to deliver the necessary proxy information to VITO for the downscaling. Therefore alternatives were sought andwe had to resort to the EDGAR v4.2 inventory for sector split information. The EDGAR v4.2 information contains global emission estimates on a 0.1x0.1° grid. These contain emissions in the sectors energy, industry, transport, residential, shipping emissions and airborne emissions (landing & take-off, climb & descent and cruising). From this emission inventory, fraction maps were derived at the resolution of the low resolution satellite imagery (0.25°) delivered by KNMI.

- *Proxy data provided by QEERI:* According to QEERI, the most important sources with respect to NOx emissions in Qatar, can be found within the following sectors:
  - $\rightarrow$  Anthropogenic sources:
    - o Oil and Gas
    - o Off shore Oil and Gas
    - Water desalination & Power generation
    - o Road Transport
    - o Shipping



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:57 of 64

- o Aviation
- Residential & Commercial : to a lesser degree
- o Agriculture, Farms?
- $\rightarrow$  Biogenic emissions : absent

Consequently, good proxy data for each of these sectors are required. Given the decision to resort to the EDGAR v4.2 inventory for sector split information (see above), proxy data for the EDGAR sectors, being energy, industry, transport, residential, shipping emissions and airborne emissions (landing & take-off, climb & descent and cruising), were sought either from open-source and freely available datasets, or was constructed using various range of GIS or classification techniques. Below is a short overview:

- Road network : <u>www.openstreetmap.org</u> was used to calculate the total length of road per gridcel (both high & low resolution)
- Energy emissions (powerplants) were estimated using the http://enipedia.tudelft.nl archive.
- The residential and industrial emissions were disaggregated using a land cover map which was derived via support vector regression classification of two landsat 8 images (dated 15 september 2015).
- Maritime and off-shore activities were disaggregated using an AIS transponder density map from http://www.marinetraffic.com
- Finally airborne emissions were disaggregated using flight tracks from www.openflights.org

Sector	Source	Proxy	Unit
Transport	Open Street Maps	Total length of roads	km
Energy	Enipedia	CO2 emissions	kton
Industry	Landsat	Built up area	4 km <sup>2</sup>
Residential	Landsat	Built up area	4 km <sup>2</sup>
Shipping	marinetraffic.com	AIS density color map	a.u.
Air - LTO	manual	Airport surface	4 km²
Air - CDS	flightdata.org	Total length of flighttracks to DOH	Km

This yields the following mapping between the sectors and the proxy data:



Below we give a brief overview of these proxy layers derived for Qatar.



# Total length of road [km] based on open street maps :

Figure 6 : Total length of roads on the coarse and high resolution grid (right) for Qatar.

#### Location of powerplants in Qatar (enipedia) :

These were derived base on the enpedia database (<u>http://enipedia.tudelft.nl/</u>)., which contains next to name, latitude, longitude also information on power output in MW, capacity and/or CO2 emissions. We are using the information on CO2 emission as a proxy to downscale the NOx emissions. Where the CO2 emission information is missing, we estimate it by scaling the power output or powerplant capacity using the CO2 emissions/MW ratio calculated from the rest of the Qatari powerplants. Below we give a table with the reconstructed information used as a proxy to downscale the power plant emissions.





#### Figure 7 : Location of powerplants in Qatar, source: enipedia (http://enipedia.tudelft.nl/)

Name	Lat.	Lon.	CO2 emissions [kton]	Capacity [MW]
Ras Abu Fontas B1	25.19841	51.61591	1549	985
Ras Laffan-a	25.92279	51.54527	1277	756
Ras Abu Fontas A	25.20497	51.6177	861	626
Ras Laffan-b	25.92462	51.54786	790	1025
Umm Said Refinery	24.92483	51.55808	365	128
Al-wajbah	25.3001	51.40455	360	301
Ras Laffan Rasgas	25.89208	51.54192	358	330
Ras Laffan Qatargas	25.90949	51.55555	206	187
Saliyah	25.20661	51.39243	164	134
Doha South Super	25.19359	51.52414	84	67
Dukhan Field	25.4228	50.75307	52	44
Mesaieed Qatalum	24.97119	51.58123	1906	1350



	-			
Mesaieed	24.97516	51.57787	2834	2007
Ras Laffan-c	25.93635	51.52407	3855	2730
Ras Abu Fontas B2	25.19632	51.61605	801	567
Umm Said Qapco	24.99611	51.54889	77	N/A
Doha Dswmc	25.28667	51.53333	N/A	N/A
Ras Laffan Dolphin	25.90861	51.57833	N/A	N/A
Al Ali Hospital	25.27932	51.52245	N/A	N/A
Doha Qafco	25.27932	51.52245	N/A	N/A
Qafco Works	24.98751	51.5466	286	N/A
Ras Abu Aboud	25.32342	51.50868	193	N/A
Mesaieed Qvc	24.98751	51.5466	153	N/A
Halul Terminal	25.66667	52.41667	16	N/A
Abu-samra	25.21667	50.96639	5	N/A
Maersk Qatar	25.35483	51.18388	25	N/A

Table 3 : Proxy information derived from enipedia to downscale the powerplant fraction of the low resolution emissions.

#### Proxy data for residential/industrial emissions :

In order to derive a proxy dataset for residential/industrial emissions, we typically rely on land cover/land use information. However for Qatar, since no dataset is available, we derived an own land cover map based upon support vector classification of a landsat image :

- two landsat 8 images (15 september 2015) were downloaded and mosaicked together to cover the full Qatari peninsula.
- Ground truth samples were constructed using "best guess" and Google Earth. We used 5 different classes in the classification : Urban (builtup-area), Vegetated area, Roads, Desert (Bare Soil) and Waterbodies. Unfortunately, within the project it was not possible to distinguish adequately between urbanised area's and industrial area's.
- The classification was performed using support vector machine in the SAGA-GIS software package.
- Afterwards, manual corrections were applied based upon google earth.

The resulting land cover map is given below :



Title:	Algorithm Theoretical Baseline
Ref.:	GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	61 of 64



Figure 8 : Left : false color image of the Landsat 8 image (15 september 2015) used for the classification, shown on the right. Below, a zoom over Doha is provided.



Title:	Algorithm Theoretical Baseline
Ref.:	GE_ATBD
Issue:	3.4
Date:	31 March 2016
Page:	62 of 64

From this land cover classification dataset, the following residential/built up proxy patterns are derived, both at low and high resolution



#### Proxy data for shipping emissions :

Shipping emissions were downscaled with information reconstructed from marine traffic transponder density maps for 2013 available from <u>http://www.marinetraffic.com</u>. These density maps clear show the presence of important oil and gas fields in the Arabian gulf. The coarse and high resolution proxy datasets are shown below.



Figure 10 : Coarse and high resolution density maps used to downscale the shipping emissions, derived from the 2013 AIS transponder density maps on marinetraffic.com



# 4.7.2 Algorithm description

Below, the methodology used for downscaling of NOx emissions for the Middle East is outlined. The methodology for downscaling the regional inventory consisted of the following steps:

- 1. The gridded regional (0.25° resolution) emission estimates from KNMI were combined with EDGAR emissions as to obtain *sector-specific emission estimates*. Hereto, from the EDGAR emission inventory, fraction maps were derived at the resolution of the low resolution satellite imagery (0.25°) delivered by KNMI. Subsequently, the EDGAR-based fractions were applied to the KNMI emission estimates. As a result, the KNMI emission estimates were split into sectoral emissions (with preservation of the KNMI total per grid cell);
- 2. Sector-specific *emission estimates resulting from step one were smoothened* as to avoid boundary effects. Hereto a bilinear interpolation was applied;
- 3. Spatial patterns were generated using the available geographical proxy data (see above);
- 4. Computation of high resolution emission maps (0.025<sup>•</sup>) from
  - a. the smoothened sectoral emissions (step 2)
  - b. the spatial patterns on the high resolution degrees grid (step 3)

# 4.7.3 Example product

An example product is shown below.



Title:Algorithm Theoretical BaselineRef.:GE\_ATBDIssue:3.4Date:31 March 2016Page:64 of 64



Figure 11 :Example product showing the downscaled emission estimates above Qatar for the month of January 2011.