

Towards automation of chemical process route selection based on data mining

P.-M. Jacob, P. Yamin, C. Perez-Storey, M. Hopgood, A. Lapkin¹

Department of Chemical Engineering and Biotechnology, University of Cambridge,

Cambridge CB2 3RA, United Kingdom

Supplementary Information

Supplementary Information includes details of calculations of exergy efficiency, prediction of molecular properties and details of network search.

Evaluation of Exergy

For an incompressible liquid of constant heat capacity Equation (3) of the main article can be expressed as:¹

$$Ex_{i,ph_{liquid}} = n_i \left(c_{p_i} (T - T_0) - T_0 c_{p_i} \ln \frac{T}{T_0} + v_i (P - P_0) \right) \quad (1)$$

where v_i is the molar volume of the substance $\left(v_i = \frac{MR}{\rho} \right)$, where MR is the molar mass and ρ the mass density of substance i . In the case of an ideal gas the physical exergy is given by:

$$Ex_{i,ph_{gas}} = n_i \left(c_{p_i} (T - T_0) - T_0 c_{p_i} \ln \frac{T}{T_0} + RT_0 \ln \frac{P}{P_0} \right) \quad (2)$$

Should the physical exergy of a substance be calculated that is a vapour at the given conditions but a liquid at the reference state the exergy change due to vaporisation needs to additionally be taken into account:

$$\Delta S_{vap} = \frac{\Delta H_{vap}}{T_{vap}} \quad (3)$$

¹ Corresponding author: aal35@cam.ac.uk. Fax: +44 1223 334796.

$$Ex_{i,ph_{vap}} = n_i \left(\Delta h_{vap} + T_0 \frac{\Delta h_{vap}}{T_{vap}} \right) \quad (4)$$

Thus, also considering the exergy changes in bringing the liquid, should it be sub-cooled, to its boiling point, T_{vap} , the following expression is obtained assuming that the liquid is boiled at P_0 :

$$Ex_{i,ph} = n_i \left(c_{p,i,liquid} (T_{vap} - T_0) - T_0 c_{p,i,liquid} \ln \frac{T_{vap}}{T_0} \right) \quad (5)$$

Due to Hess's Law it does not matter which path is taken to reach T, P from T_0, P_0 , care however must be taken that should vaporisation not be carried out at atmospheric pressure the correct values for Δh_{vap} and T_{vap} are used.¹

Process Heating

Electrical energy, unlike other fuel media, can be fully converted into work and thus the exergy content of electricity is equal to the energy content of the work done by it:¹⁻⁴

$$Ex_{Q_{elec}} = Q \quad (6)$$

Furthermore, the model has to account for the efficiency of electricity production. Firstly there are losses in transmission and distribution⁵ which amount to approximately 1.5% and 5%, respectively.⁶ Values on the exergetic efficiency of power production from fossil fuels in the UK quoted in literature vary between 32.1-33.5%,⁵ 33%,⁷ 37%,⁸ and 45%.⁹ Taking the average of these values yields an efficiency of 37%. Furthermore, it is assumed that the heat exchanger employed is, on average, 80% efficient. Thus:

$$Ex_{elec} = \frac{Q}{\eta_{HEX} \eta_{transmission \text{ and distribution}} \eta_{generation}} \quad (7)$$

$$Ex_{elec} = 3.57Q \quad (8)$$

Exergetic Efficiency

The exergetic efficiency of a process is often defined as the ratio of useful exergy output to exergy input as given by Equation (9).

$$\eta = \frac{Ex_{out}}{Ex_{in}} \quad (9)$$

This definition however suffers from a drawback: solvents and unreacted starting material will appear in the process output without actually having been "produced" by the process thus artificially inflating the efficiency.¹⁰ Instead a more useful approach would be to only consider the produced, utilisable exergy in relation to the consumed exergy, thus excluding the transiting exergy associated to the parts of the mixture not taking part in the reaction as shown in Equation (10) where Ex_{tr} is the transiting exergy given by the sum of the transiting exergy across all substances i as given in Equation (11) and (12) where $Ex_{i,in}$ and $Ex_{i,out}$ are the exergies associated to species i entering and exiting the system, respectively.¹⁰

$$\eta = \frac{Ex_{pu}}{Ex_c} = \frac{Ex_{out} - Ex_{tr}}{Ex_{in} - Ex_{tr}} \quad (10)$$

$$Ex_{tr} = \sum_i Ex_{tr,i} \quad (11)$$

$$Ex_{tr,i} = n_{i,tr} \min(ex_{i,ch,in}, ex_{i,ch,out}) + n_{i,tr} \text{mir} \quad (12)$$

Mass-based indicators

According to Andraos's paper:¹¹

$$E_{total} = E_{byproducts\&unreacted\ reagents} + E_{excess\ reagents} + E_{auxiliaries} \quad (13)$$

$$E_{\text{byproducts\&unreacted reagents}} = \frac{1}{MR_{p_n}} \sum_j \left(\frac{1}{\prod_k^{\text{n}\rightarrow\text{j}} \varepsilon_k} \frac{MR_{p_j}}{AE_j} [1 - \varepsilon_j AE_j] \right) \quad (14)$$

$$E_{\text{excess reagents}} = \frac{1}{MR_{p_n}} \sum_j \left(\frac{1}{\prod_k^{\text{n}\rightarrow\text{j}} \varepsilon_k} \frac{MR_{p_j}}{AE_j} [SF_j - 1] \right) \quad (15)$$

$$E_{\text{auxiliaries}} = \frac{1}{MR_{p_n}} \sum_j \left(\frac{1}{\prod_k^{\text{n}\rightarrow\text{j}} \varepsilon_k} \frac{c_j + s_j + \omega_j}{n_{r_j}} \right) \quad (16)$$

where E is the E-factor, MR_p the molecular weight of the product, ε is the yield with respect to the limiting reactant. The subscripts j and n relate to reaction step number j in the synthesis route and the final step, respectively, where the sequence of steps is $(1, \dots, j, \dots, n)$.

$\prod_k^{\text{n}\rightarrow\text{j}} \varepsilon_k$ is the product of reaction yields along the reaction route from the current step to the final step, ignoring any steps carried out prior to the current step, c is the mass of catalyst, s is the mass of solvent, ω is the mass of all other materials used in work-up and purification and n_{r_j} is the experimental mole scale of the limiting reagent in step j . The subsequent derivations assume that the feed contains only reactant, reagents and auxiliaries, i.e. no products or byproducts and thus the yield is being defined as:

$$\varepsilon = \frac{n_p}{n_{mr}} \quad (17)$$

Andraos defines the atom economy, AE_j , and the stoichiometric factor, SF_j , as:

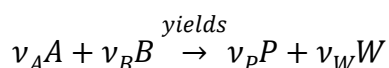
$$AE_j = \frac{1}{1 + (E_{mw})_j} \quad (18)$$

$$SF_j = 1 + \frac{\sum \text{excess mass of all reagents}}{\sum \text{stoichiometric masses of all reagents}} \quad (19)$$

where $(E_{mw})_j$ is the E-factor of step j based on the molecular weights:

$$(E_{mw})_j = \frac{\sum MR_{\text{byproducts}}}{MR_{p_j}} \quad (20)$$

This definition however only works for reactions where all stoichiometric coefficients are one and thus the equations given in¹¹ do not hold for more complex systems unless modified. All modifications are based on the following reaction where A is the limiting reactant, B a reagent, P the product and W a byproduct:



Modifying Equation (20) to take stoichiometry into account we obtain:

$$E = \frac{\sum_{bp} \nu_{bp} MR_{bp}}{\nu_p MR_p} \quad (21)$$

where ν is the stoichiometric coefficient and bp refers to the biproducts, thus under this definition there only ever is one product, and MR is the molecular weight of a species. The product of stoichiometric coefficient and molecular weight shall be defined as stoichiometric weight, SW , henceforth. Substituting Equation (21) into (18) we obtain:

$$AE_j = \frac{1}{1 + \left(\frac{SW_{bp,j}}{\nu_{p,j} MR_{p,j}} \right)} \quad (22)$$

$$AE_j = \frac{1}{\frac{SW_{p,j} + SW_{bp,j}}{v_{p,j}MR_{p,j}}} \quad (23)$$

Noting however that due to the conservation of mass the stoichiometric mass of reactants ($SW_{r,j}$) equals to that of the products and byproducts the following expression is obtained:

$$AE_j = \frac{v_p MR_{p,j}}{SW_{r,j}} \quad (24)$$

At this point we define the subscript mr to denote the main reactant. Usually this is the starting material of the reaction and thus normally also the limiting reactant. Seeing as many equations will be normalised with respect to this care must be taken when the main reactant is not the limiting reactant to ensure the scaling is still correct.

The mass of byproducts formed and unreacted reagents, m_{bp} is given by the difference in mass of stoichiometric amount of reactants and mass of product formed as shown:

$$m_{bp} = n_{mr} \frac{SW_r}{v_{mr}} - n_p MR_p \quad (25)$$

This is in turn normalised with respect to the number of moles of starting material used to allow scaling, yielding m_{bp}^* :

$$m_{bp}^* = \frac{n_{mr} \frac{SW_r}{v_{mr}} - n_p MR_p}{n_{mr}} \quad (26)$$

Carrying out rearrangements of Equation (26) we obtain the following sequence of steps:

$$m_{bp}^* = SW_r \frac{n_{mr} SW_r - v_{mr} n_p MR_p}{v_{mr} n_{mr} SW_r}$$

$$m_{bp}^* = SW_r \left(\frac{1}{v_{mr}} - \frac{v_{mr} n_p MR_p}{v_{mr} n_{mr} SW_r} \right)$$

$$m_{bp}^* = SW_r \left(\frac{1}{v_{mr}} - \frac{v_{mr} n_p v_p MR_p}{v_p v_{mr} n_{mr} SW_r} \right)$$

$$m_{bp}^* = \frac{SW_r}{v_{mr}} \left(1 - \frac{v_{mr} n_p v_p MR_p}{v_p n_{mr} SW_r} \right)$$

Recalling Equations (24) and (17) we obtain:

$$m_{bp}^* = \frac{v_p MR_p}{v_{mr} AE} \left(1 - \frac{v_{mr}}{v_p} \varepsilon AE \right) \quad (27)$$

Adjusting Equation (27) for the cumulative yield is possible to calculate the E-factor:

$$E_{byproducts\&unreacted\ reagents} = \frac{1}{MR_{p_n}} \sum_j \left(\frac{1}{\prod_k \varepsilon_k} \left(\frac{v_{p_j} MR_{p_j}}{v_{mr_j} AE_j} \left[1 - \frac{v_{mr_j}}{v_{p_j}} \varepsilon_j AE_j \right] \right) \right) \quad (28)$$

The expression for the mass of excess reagents can be similarly derived. The normalised

mass of excess reagents, m_{excess} , is given by:

$$m_{excess}^* = \frac{m_{excess}}{n_{mr}} \quad (29)$$

Equation (29) can in turn be rearranged as follows:

$$m_{excess}^* = \frac{SW_r}{v_{mr}} \left(\frac{m_{excess}}{n_{mr} \frac{SW_r}{v_{mr}}} \right)$$

$$m_{excess}^* = \frac{v_p MR_p}{v_{mr} AE} \left(\frac{m_{excess}}{n_{mr} \frac{SW_r}{v_{mr}}} \right)$$

Equation (19) can be expanded into a more rigorous form:

$$SF = 1 + \frac{m_{excess}}{n_{mr} \frac{SW_r}{v_{mr}}} \quad (30)$$

where $m_{excess} = m_{in} - n_{mr} \frac{SW_r}{v_{mr}}$. Using Equation (30) we thus obtain:

$$m_{excess}^* = \frac{v_p MR_p}{v_{mr} AE} (SF - 1) \quad (31)$$

Adjusting this for the cumulative yield across a route the following E-factor expression is obtained:

$$E_{excess\ reagents} = \frac{1}{MR_{p_n}} \sum_j \left(\frac{1}{\prod_k \varepsilon_k} \left(\frac{v_{p_j} MR_{p_j}}{v_{mr_j} AE_j} [SF_j - 1] \right) \right) \quad (32)$$

Equation (16) holds unchanged. Thus, combining Equations (16), (28) and (32) an overall expression for the total E-factor across a route can be obtained:

$$E_{total} = \frac{1}{MR_{p_n}} \sum_j \left(\frac{1}{\prod_k \varepsilon_k} \left(\frac{v_{p_j} MR_{p_j}}{v_{r_j} AE_j} \left[SF_j - \frac{v_{mr_j}}{v_{p_j}} \varepsilon_j AE_j \right] + \frac{c_j + s_j + \omega_j}{n_{mr_j}} \right) \right) \quad (33)$$

Property Prediction

Table S1. Liquid heat capacity predictions.

Species	c_p [J mol ⁻¹ K ⁻¹]
alpha-Terpinene	251.4
gamma-Terpinene	251.4
alpha-Terpinolene	250.3
p-Propyltoluene	246.4
4-Carboxybenzaldehyde	244.9

Table S2. Gas heat capacity predictions.

Species	c_p [J mol ⁻¹ K ⁻¹]
p-Cymene(g)	330.21
o-Cymene(g)	321.77
m-Cymene(g)	329.96
p-Propyltoluene(g)	394.33

Table S3. Gibb's energy of formation predictions.

Species	ΔG_f° [kJ mol ⁻¹]
alpha-Terpinene	103.7
gamma-Terpinene	103.7
alpha-Terpinolene	122.72
p-Ethyltoluene	127.68
p-Propyltoluene	136.1
Terephthalic acid	-568.56
p-Toluic acid	-224.65
4-Carboxybenzaldehyde	-324.17

Table S4. Heat of reaction prediction.

Species	ΔG_f° [kJ mol ⁻¹]	$v_{1,1}$	$v_{1,2}$	$v_{1,3}$	$v_{2,1}$	$v_{2,2}$	$v_{3,1}$	$v_{3,2}$	v_4	v_5
Limonene	-54.5	-1	-1	-1						
p-Cymene	78	0.32	0.88	1						
p-Cymene (gas)	127.5				-1	-1			-1	
o-Cymene (gas)	122.2				0.003					
m-cymene (gas)	125.9				0.108					
toluene(gas)	49.4				0.866	0.96				
toluene (liquid)	12.4						-1	-1		
propylene (gas)	20				0.867	0.75				
alpha-terpinene	-236	0.25	0.09							
gamma-terpinene	-236	0.18	0.03							
alpha-terpinolene	-113.17	0.25								
p-xylene (gas)	17.9				0.001					
benzene (gas)	82.1				0.001					
p-propyltoluene (gas)	-25.35				0.021					
hydrogen peroxide	-187.8						-0.91	-1		
oxygen	0							-1	-5.34	-0.31
benzoic acid	-385.06						0.82	1		0.25
benzyl alcohol	-160.7						0.18			
water	-285.83							2	3.56	
methane	-74.6				0.003					
hydrogen	0	0.32	0.88	1	-0.002		1.73			0.37
terephthalic acid	-816.3								0.56	
p-toluic acid (l)	-404.51								0.44	-1
carbon dioxide	-393.52								2	0.25
4-carboxybenzaldehyde	-479.5									0.12
ΔH_R (kJ/mol)		-50.31	94.82	132.50	-54.07	-65.08	-186.18	-781.32	-2567.21	152.33

Table S5. Gibb's free energy of mixing for the species contained in the outlet stream of reaction 1.1.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
Limonene	0.02399	0.725	-3.70E-02
alpha-Terpinine	0.00172	0.052	-7.43E-03
gamma-Terpinene	0.00124	0.037	-5.75E-03
alpha-Terpinolene	0.00172	0.052	-7.40E-03
p-Cymene	0.00220	0.067	-8.77E-03
Hydrogen	0.00220	0.067	-9.69E-03

Table S6. Gibb's free energy of mixing for the species contained in the outlet stream of reaction 1.2.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
Limonene	0.00000	0.000	0
alpha-Terpinine	0.00278	0.048	-1.22E-02
gamma-Terpinene	0.00093	0.016	-5.09E-03
alpha-Terpinolene	0.00000	0.000	0
p-Cymene	0.02717	0.468	-5.52E-02
Hydrogen	0.02717	0.468	0

Table S7. Gibb's free energy of mixing for the species contained in the outlet stream of reaction 1.3.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
Limonene	0.00000	0	0
p-Cymene	0.03087	0.5	-4.29E-02
Hydrogen	0.03087	0.5	0

Table S8. Gibb's free energy of mixing for the species contained in the outlet stream of reaction 2.1.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
p-Cymene	9.400E-08	1.693E-05	-5.78E-03
Benzene	3.060E-10	5.513E-08	-2.86E-05
Toluene	2.650E-07	4.774E-05	-1.48E-02
p-Xylene	3.060E-10	5.513E-08	-2.86E-05
p-Propyltoluene	6.426E-09	1.158E-06	-4.91E-04
o-Cymene	9.180E-10	1.654E-07	-8.02E-05
m-Cymene	3.305E-08	5.954E-06	-2.22E-03
Methane	9.180E+00	1.654E-07	-8.02E-05
Hydrogen	5.550E-03	9.990E-01	-3.73E-03
Propylene	2.653E-07	4.780E-05	-1.48E-02

Table S9. Gibb's free energy of mixing for the species contained in the outlet stream of reaction 2.2.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
p-Cymene	5.960E-04	0.0547	-10.41
Toluene	5.782E-03	0.531	-22.02
Propylene	4.517E-03	0.415	-23.91

Table S10. Gibb's free energy of mixing for the species contained in the outlet stream of reaction 3.1.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
Toluene	0.00075	0.001	3.35E-03
Hydrogen peroxide	0.45413	0.698	-7.43E-01
Benzoic acid	0.00349	0.005	-1.96E-03
Benzyl alcohol	0.00077	0.001	-2.58E-03
Acetonitrile	0.19171	0.295	-7.45E-01

Table S11. Gibb's free energy of mixing for the species contained in the outlet stream of reaction 3.2.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
Toluene	0.00080	0.002	-1.31E-03
Hydrogen peroxide	0.02280	0.064	-1.36E-01
Oxygen	0.00000	0.000	
Benzoic acid	0.01520	0.043	-8.00E-02
Water	0.03040	0.085	-1.21E-02
Acetonitrile	0.28756	0.806	-5.98E-02

Table S12. Gibb's free energy of mixing for the gaseous species contained in the outlet stream of reaction 4.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
p-Cymene	0.293	0.0017	-7376.35
Air	164.341	0.9303	-46710.27
			-
Water	9.388	0.0531	108345.95

Table S13. Gibb's free energy of mixing for the gaseous species contained in the outlet stream of reaction 5.

Species	Out [mol]	x_{out}	ΔG_{mix} [kJ]
p-Toluic acid	0.00903	6.377E-04	-1.25E-01
4-Carboxybenzaldehyde	0.00018	1.246E-05	-3.00E-03
Benzoic acid	0.00037	2.595E-05	-5.82E-03
Oxygen	0.00000	0.000E+00	0.00E+00
Hydrogen	0.00054	3.841E-05	-5.30E-03
Carbondioxide	0.00037	2.595E-05	-3.86E-03
Trifluoroacetic acid	12.73500	8.993E-01	-13.19
Water	1.41500	9.993E-02	-13.30

Network Search

Table S14. Overview of the possible three step synthesis routes, as retrieved from Reaxys, connecting limonene to benzoic acid according to the product of the first synthesis step, ranked in decreasing order of occurrence. The table lists the number of routes going via a given product.

Species	Number of Occurrences
Cumol	60
Maleic anhydride	38
p-Cymene	29
Fumaric acid	21
Methyl 4-methylphenyl ketone	19
Thymol	8
dimethyl Terephthalate	7
4-Ethenylcyclohexene	7
1-Methyl-4-isopropenylbenzene	5
alpha-Terpinene	4
1,8-Epoxy-p-menthane	4
(R)-(+)-limonene	3
4-methylcumyl alcohol	2
Terpinolene	2
2-(4'-Methylphenyl)propionaldehyde	2
1,8-Dichloro-p-menthane	1
Dihydrocarvone	1
2,2,6-Trimethyl-4-phenyl-3-oxabicyclo<3.3.1>non-6-ene	1

4-Methyl-1-isopropylidenecyclohexane	1
4-Acetyl-1-methyl-cyclohex-1-ene	1
Terpin	1
8,9-Epoxy-p-menth-1,2-ene	1
Piperitenone	1
6-Isopropyl-3-methyl-2-cyclohexen-1-one	1
p-Mentha-1,5-diene	1
5-Isopropyl-2-methylcyclohexanol	1
cis-p-Menthane	1
4-Isopropylbenzoic acid	1

Exergetic Efficiency Results

Table S15. Overview of exergetic efficiencies of the reactions. *Ex* denotes exergy and the subscripts *in*, *HEX*, *HR*, *Gm*, *out* and *tr* indicate that this is the exergy associated to the streams entering the system, heating of inlet streams, reaction heating, the Gibb's Free Energy of Mixing required for separation, a species leaving the system and a species transiting the system, respectively. η is the exergetic efficiency in percent. All exergies are given as *kJ*.

	1.1	1.2	1.3	2.1	2.2	3.1	3.2	4	5
EX _{in}	189.84	189.84	189.84	3.64	38.98	321.34	428.28	17565.04	9025.52
EX _{HEX}	3.87	2.77	4.29	0.45	4.26	8.52	8.20	4463.18	106.10
EX _{HR}	0.00	10.45	14.60	0	0.00	0	0.00	0.00	0.00
EX _{Gm}	0.08	0.07	0.00	16.35	56.33	1.49	0.29	162432.57	26.64
EX _{prod}	12.99	160.09	182.05	1.07	23.28	11.65	50.85	4642.23	1.23
EX _{reactants,out}	147.66	0	0	0.56	3.57	64.13	6.23	1603.86	36.17
EX _{reactants,tr}	147.50	0	0	0.55	3.51	64.02	6.23	1603.86	36.17
EX _{solvents,out}	0	0	0	1.31	0	240.11	360.30	-51540.00	8984.27
EX _{solvents,tr}	0	0	0	1.28	0	240.06	360.09	-51749.03	8983.46
η	28.41%	78.81%	87.22%	5.90%	24.30%	43.31%	72.48%	2.07%	1.47%

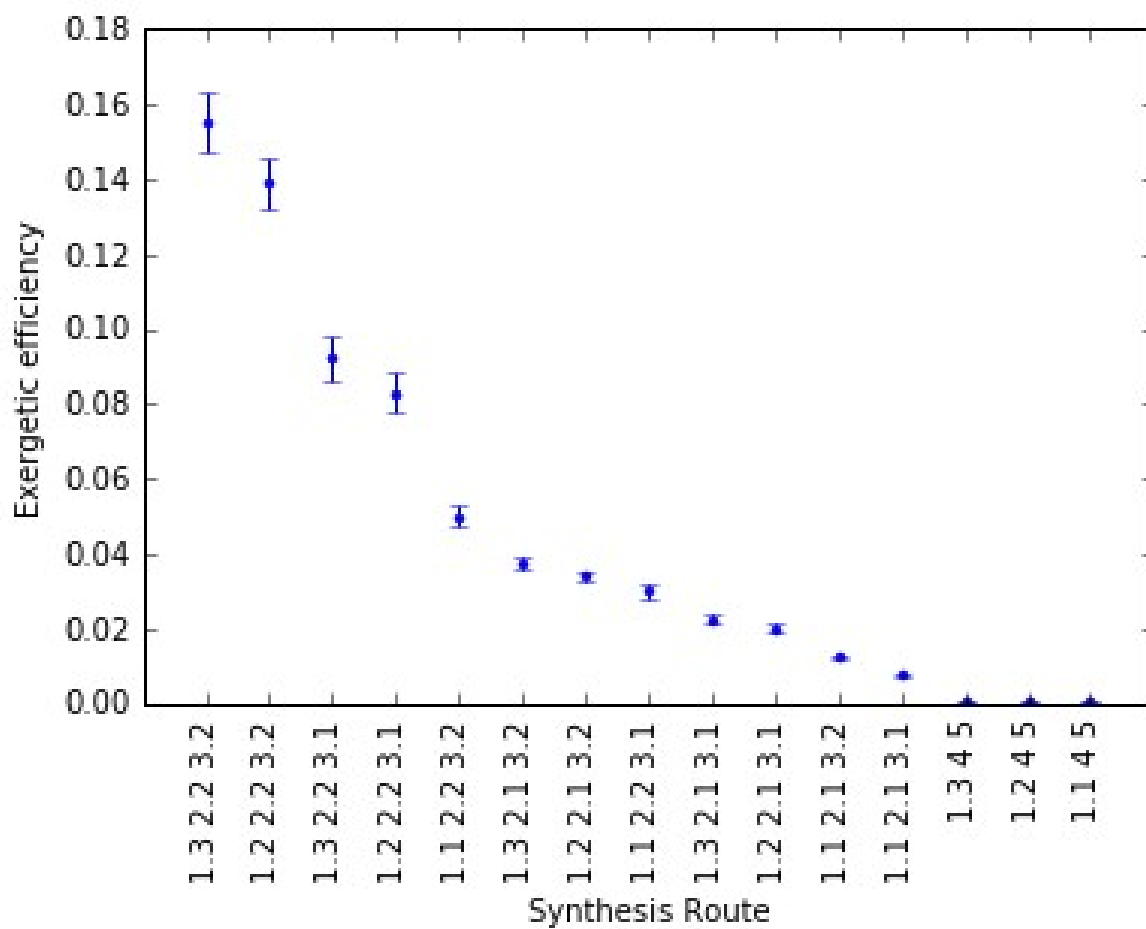


Figure S1: Plot of the exergetic efficiency for each synthesis route using linear error propagation as implemented in reference 12.

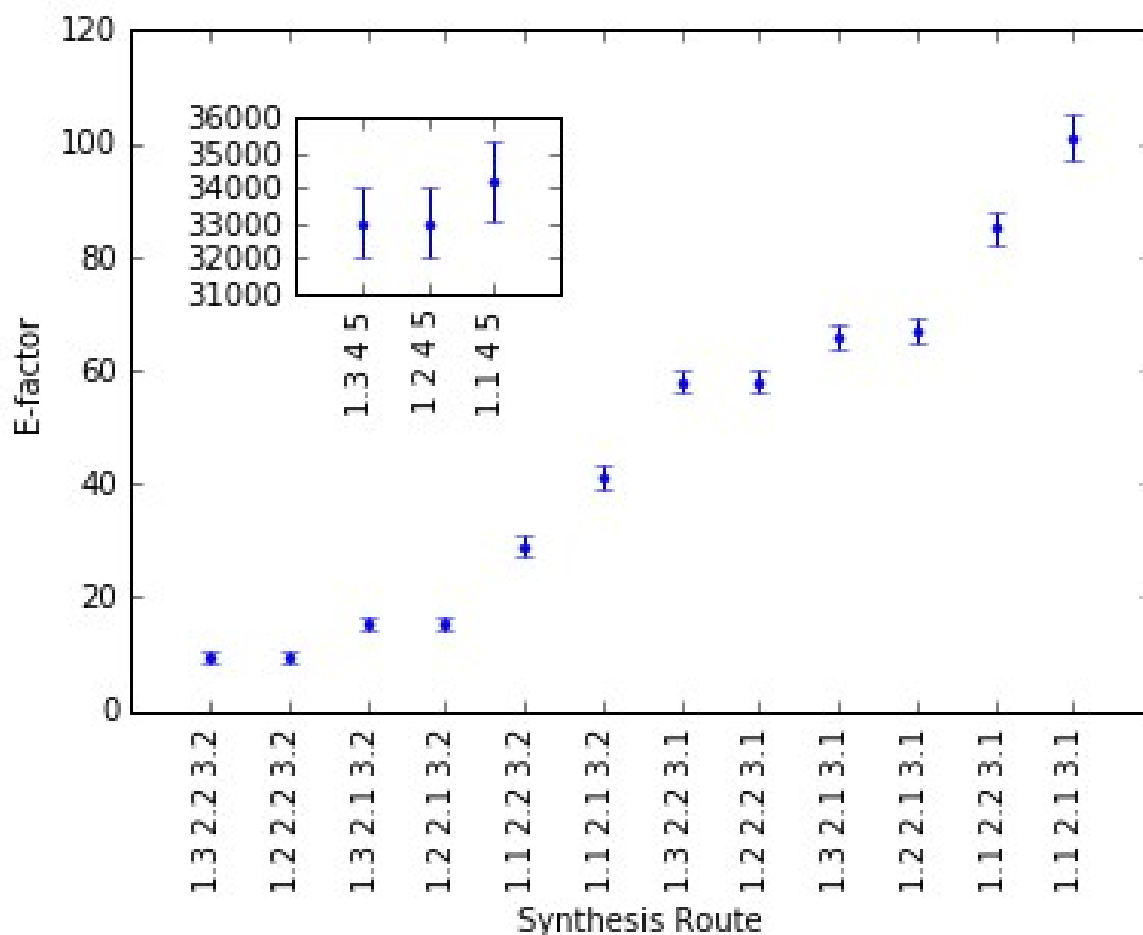


Figure S2: Plot of the E-factor for each synthesis route using linear error propagation as implemented in reference 12.

Bibliography

- 1 E. Querol, B. Gonzalez-Regueral and J. L. Perez-Benedito, in *Practical Approach to Exergy and Thermo-economic Analyses of Industrial Processes*, Springer London, London, 2013, pp. 9–28.
- 2 M. A. Rosen and I. Dincer, *Energy Convers. Manag.*, 2003, **44**, 1633–1651.
- 3 J. Dewulf, H. Van Langenhove, B. Muys, S. Bruers, B. R. Bakshi, G. F. Grubb, D. M. Paulus and E. Sciubba, *Environ. Sci. Technol.*, 2008, **42**, 2221–2232.
- 4 R. U. Ayres, L. W. Ayres and K. Martinas, *Eco-Thermodynamics: Exergy and Life Cycle Analysis*, Fontainebleau, 1996.

- 5 P. J. Boait, R. M. Rylatt and A. Wright, *Appl. Energy*, 2007, **84**, 239–253.
- 6 H. Alderson, G. R. Cranston and G. P. Hammond, *Energy*, 2012, **48**, 96–107.
- 7 G. P. Hammond and A. J. Stapleton, *Proc. Inst. Mech. Eng. Part A J. Power Energy*, 2001, **215**, 141–162.
- 8 M. A. Rosen and C. A. Bulucea, *Entropy*, 2009, **11**, 820–835.
- 9 G. P. Hammond, *Energy Policy*, 2004, **32**, 1789–1798.
- 10 M. Sorin, J. Lambert and J. Paris, *Chem. Eng. Res. Des.*, 1998, **76**, 389–395.
- 11 J. Andraos, *Org. Process Res. Dev.*, 2009, **13**, 161–185.
- 12 E. O. Lebigot, Uncertainties: a Python package for calculations with uncertainties, <http://pythonhosted.org/uncertainties/>.