Supporting Information

Automated versus chemically intuitive deconvolution of DFT-based gas-phase errors in nitrogen compounds

Ricardo Urrego-Ortiz,¹ Santiago Builes,¹,* Federico Calle-Vallejo²,³,4,*

¹ Departamento de Ingeniería de Procesos, Universidad EAFIT, Carrera 49 # 7 sur 50, 050022, Medellín, Colombia.

² Department of Materials Science and Chemical Physics & Institute of Theoretical and Computational Chemistry, University of Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain.

³ Nano-Bio Spectroscopy Group and European Theoretical Spectroscopy Facility (ETSF), Department of Polymers and Advanced Materials: Physics, Chemistry and Technology, University of the Basque Country UPV/EHU, Avenida Tolosa 72, 20018 San Sebastián, Spain.

⁴ IKERBASQUE, Basque Foundation for Science, Plaza de Euskadi 5, 48009 Bilbao, Spain.

*Correspondence to: sbuiles@eafit.edu.co, f.calle.vallejo@ub.edu

Table of contents

S1. Schematics of the nitrogen compounds under study 2
S2. Tabulated data and convergence tests 2
S3. Matrix representations of the nitrogen compounds 4
S4. Optimization procedure 5
S5. Free and fixed parameter tests and comparison to previous works 6
S6. Enthalpic effects 9
S7. Analysis of the initial trends in Figure 5 10
S8. Direct coordinates 10
References 11
S1. Schematics of the nitrogen compounds

Figure S1 shows the ball-and-stick representations, the names and molecular formulas of the 13 nitrogen compounds under study.

![Schematics of nitrogen compounds](image)

**Figure S1.** Ball-and-stick representation of the nitrogen compounds in this study. Blue balls correspond to nitrogen atoms, red balls to oxygen atoms and white balls to hydrogen atoms.

S2. Tabulated data and convergence tests

Table S1 shows the DFT-calculated zero-point energies (ZPEs) and TS corrections at T = 298.15 K for the 13 compounds under study together with the values for H₂, O₂, N₂, H₂O and NH₃, which are necessary for the assessment of the errors.

Figure S2 shows that the formation energies of N₂O are converged at a plane-wave cutoff of 450 eV. The convergence of the ZPEs is illustrated in Table S2.
Table S1. DFT-calculated ZPE and experimental\textsuperscript{1,2} TS corrections for the compounds in this study at 298.15 K. All values are in eV and were calculated with PBE.

| species       | ZPE  | TS  |
|---------------|------|-----|
| H\textsubscript{2}    | 0.27 | 0.40 |
| O\textsubscript{2}     | 0.10 | 0.63 |
| N\textsubscript{2}     | 0.15 | 0.59 |
| H\textsubscript{2}O   | 0.57 | 0.58 |
| NH\textsubscript{3}   | 0.91 | 0.60 |
| NH\textsubscript{2}OH | 1.06 | 0.73 |
| NO                | 0.12 | 0.65 |
| HNO               | 0.37 | 0.68 |
| NO\textsubscript{2}  | 0.23 | 0.74 |
| NO\textsubscript{3}  | 0.35 | 0.78 |
| trans-HNO\textsubscript{2} | 0.53 | 0.77 |
| cis-HNO\textsubscript{2} | 0.53 | 0.77 |
| HNO\textsubscript{3} | 0.69 | 0.82 |
| N\textsubscript{2}O   | 0.30 | 0.68 |
| cis-N\textsubscript{2}O\textsubscript{2} | 0.31 | 0.92 |
| N\textsubscript{2}O\textsubscript{3} | 0.45 | 0.97 |
| N\textsubscript{2}O\textsubscript{4} | 0.61 | 0.94 |
| N\textsubscript{2}O\textsubscript{5} | 0.71 | 1.07 |

Figure S2. Formation energy of N\textsubscript{2}O at various cutoffs. The red line is the experimental value (1.07 eV).
Table S2. Convergence of the zero-point energies. All values are in eV.

| Plane-wave cutoff | ZPE_{O2} | ZPE_{N2} | ZPE_{N2O} | ΔZPE |
|-------------------|----------|----------|-----------|------|
| 400               | 0.10     | 0.16     | 0.30      | 0.10 |
| 700               | 0.10     | 0.15     | 0.30      | 0.10 |
| 1000              | 0.10     | 0.15     | 0.30      | 0.10 |

S3. Matrix representations of the nitrogen compounds

Table S3 contains the matrix representation of the 13 compounds under study for the sequential method and AO2. A negative number is used for cis-N₂O₂ to avoid the double counting of its NN bond.

Table S3. Matrix representation of nitrogen compounds according to the sequential method and AO2.

| species | ONO | NNO | NOH | NN |
|---------|-----|-----|-----|----|
| NH₂OH   | 0.0 | 0.0 | 1.0 | 0.0|
| NO      | 0.0 | 0.0 | 0.0 | 0.0|
| HNO     | 0.0 | 0.0 | 0.0 | 0.0|
| NO₂     | 1.0 | 0.0 | 0.0 | 0.0|
| NO₃     | 1.5 | 0.0 | 0.0 | 0.0|
| trans-HNO₂ | 0.5 | 0.0 | 1.0 | 0.0|
| cis-HNO₂ | 0.5 | 0.0 | 1.0 | 0.0|
| HNO₃    | 1.0 | 0.0 | 1.0 | 0.0|
| N₂O     | 0.0 | 1.0 | 0.0 | 0.0|
| cis-N₂O₂ | 0.0 | 2.0 | 0.0 | -1.0|
| N₂O₃    | 1.0 | 1.0 | 0.0 | 0.0|
| N₂O₄    | 2.0 | 0.0 | 0.0 | 1.0|
| N₂O₅    | 2.0 | 0.0 | 2.0 | 0.0|

Table S4 contains the matrix representation of the nitrogen-containing compounds for AO1.
Table S4. Matrix representation of nitrogen compounds according to AO1.

| species      | N-O | N=O | N-N | O-H |
|--------------|-----|-----|-----|-----|
| NH$_2$OH     | 1   | 0   | 0   | 1   |
| NO           | 0   | 1   | 0   | 0   |
| HNO          | 0   | 1   | 0   | 0   |
| NO$_2$       | 1   | 1   | 0   | 0   |
| NO$_3$       | 2   | 1   | 0   | 0   |
| trans-HNO$_2$| 1   | 1   | 0   | 1   |
| cis-HNO$_2$  | 1   | 1   | 0   | 1   |
| HNO$_3$      | 2   | 1   | 0   | 1   |
| N$_2$O       | 0   | 1   | 1   | 0   |
| cis-N$_2$O$_2$| 0  | 2   | 1   | 0   |
| N$_2$O$_3$   | 1   | 2   | 1   | 0   |
| N$_2$O$_4$   | 2   | 2   | 1   | 0   |
| N$_2$O$_5$   | 4   | 2   | 0   | 0   |

S4. Optimization procedure

The nonlinear programming (NLP) optimization problem was formulated as a minimization of the objective functions:

$$MAE = \text{abs} \left( \frac{1}{n} \sum_{i=1}^{n} (\Delta_f G_i^{\text{corrected}} - \Delta_f G_i^{\text{exp}}) \right)$$  \hspace{1cm} (S1)

$$MAX = \max \left( \text{abs} (\Delta_f G_i^{\text{corrected}} - \Delta_f G_i^{\text{exp}}) \right)$$  \hspace{1cm} (S2)

where $\text{abs}()$ is the absolute value of the function, $\max()$ is the maximum value of the function and $\Delta_f G_i^{\text{corrected}} - \Delta_f G_i^{\text{exp}}$ is the absolute difference between the experimental formation Gibbs energy and the DFT corrected value for each substance $i$. A reformulation of the $\text{abs}()$ and $\max()$ functions was used to avoid the use of non-smooth functions in the optimization. The absolute value was split into the sum of its positive and negative parts, each represented by a positive variable. Thus, the discontinuity in the derivative of the function $\text{abs}()$ is converted into additional restrictions to the optimization. The max function was replaced by 13 inequalities (one for each nitrogen compound) such that the feasible space was enlarged but the discontinuity of the function was removed.

The multi-objective optimization problem formulated for AO1 found that the utopia point was within the feasible space. In other words, the minimum possible MAE
corresponds to the minimum possible MAX. However, for AO2 the utopia point was not inside the feasible space and a Pareto front was obtained, see Figure S3.

![Figure S3. Pareto front (white circles) and utopia point (red square) for AO2.](image)

**S5. Free and fixed parameter tests and comparison to previous works**

In general, it is assumed that O-H bonds are reasonably well described, but this might vary from one family of compounds to the next. AO1 is not biased by design to assume a good description of O-H bonds, but one could easily assume that error to be zero and reoptimize with it as a fixed parameter. In doing so, via AO1 we obtained a slightly larger MAE and nearly the same values for the rest of the free parameters, see Table S5. While this result calls for a thorough parameter sensitivity analysis that escapes our current scope, it suggests that there is a hierarchy of parameters in AO1 in which O-H bonds are not as important other bonds.
**Table S5.** Errors for AO1 obtained by setting the error in O-H bonds to zero (middle column) and by using it as a free parameter (right column).

| bond error | results with fixed OH / eV | original results / eV |
|------------|---------------------------|-----------------------|
| N-H        | 0.00                      | 0.00                  |
| N-O        | -0.41                     | -0.42                 |
| N=N        | -0.30                     | -0.29                 |
| N-N        | -0.23                     | -0.22                 |
| O-H        | 0.00                      | 0.18                  |
| MAE        | 0.16                      | 0.11                  |
| MAX        | 0.26                      | 0.26                  |

In addition, we set the errors in NH bonds as a free parameter in AO2 for the Pareto optimal point. Table S6 shows that the NH error is nearly zero, while the MAE, MAX and the other group errors are similar. This suggests that NH bonds are indeed generally well described.

**Table S6.** Errors for AO2 obtained by allowing the error in N-H bonds to be a free parameter (middle column) and by fixing it to be zero (right column).

| bond & group errors | results with free NH / eV | original results / eV |
|---------------------|---------------------------|-----------------------|
| NH                  | -0.02                     | 0.00                  |
| ONO                 | -0.87                     | -0.87                 |
| NNO                 | -0.42                     | -0.42                 |
| NOH                 | -0.09                     | -0.12                 |
| N-N                 | -0.01                     | -0.01                 |
| MAE                 | 0.04                      | 0.05                  |
| MAX                 | 0.08                      | 0.08                  |

In an earlier work\(^4\) we used the number of oxygen atoms \(n_o\) as a descriptor to predict the errors in the DFT-energies of 11 gaseous nitrogen compounds split in three groups, namely NO\(_x\), N\(_2\)O\(_x\) and HNO\(_x\). The rationale behind this descriptor is that the progressive addition of oxygen atoms leads to the formation of multiple bonds, which are poorly described by GGAs.\(^5\) Such correction scheme is based on Equation S3:

\[
\Delta_f H_{H_xN_yO_z}^{corr} = \left( \Delta_f H_{H_xN_yO_z}^{DFT} + \frac{y}{2} \varepsilon_{N_2} + \frac{z}{2} \varepsilon_{O_2} \right) - m_i \cdot n_o - b_i \tag{S3}
\]

Table S7 shows the errors in \(\varepsilon_{N_2}\) and \(\varepsilon_{O_2}\), together with the parameters \(m_i\) and \(b_i\) for NO\(_x\), N\(_2\)O\(_x\), and HNO\(_x\), for the PBE exchange-correlation functional.
Table S7. Parameters for the correction of errors using as a descriptor the number of oxygen atoms in the molecules. $\varepsilon_{O_2}$, $\varepsilon_{N_2}$ and $b_i$ are in eV and $m_i$ is in eV/O atom.

| Family      | parameters | value for PBE |
|-------------|------------|---------------|
| Diatomic    | $\varepsilon_{O_2}$ | -0.46         |
| Molecules   | $\varepsilon_{N_2}$ | 0.34          |
| $N_2O_x$    | $m_i$      | -0.42         |
|             | $b_i$      | 0.00          |
| $HNO_x$     | $m_i$      | -0.45         |
|             | $b_i$      | 0.39          |
| $NO_x$      | $m_i$      | -0.67         |
|             | $b_i$      | 0.59          |

Table S8 summarizes the residual DFT errors after applying the correction schemes of each study to the respective molecules using PBE. According to Table S8, all the correction schemes yield MAEs that approach chemical accuracy (~1 kcal/mol). Furthermore, the oxygen-counting procedure\textsuperscript{4} renders comparable or lower MAE and MAX values than the sequential method and AO1. The higher MAX of the oxygen-based method against AO2 probably stems from $n_o$-based corrections being unable to detect the concrete chemical structures introducing the deviations.

Table S8. Residual errors obtained after applying the correction schemes in this study and in an earlier work.\textsuperscript{4} Columns 2 to 4: the three approaches proposed here: the sequential approach ($seq$), the optimization based on chemical bonds (AO1), and the optimization that capitalizes on the errors found by the sequential approach (AO2). The last column refers to a previous correction scheme.\textsuperscript{4} NA: not analyzed.

| species   | this work | earlier work |
|-----------|-----------|--------------|
|           | $seq$    | $AO1$ | $AO2$ | $n_o$         |
| NH$_2$OH  | 0.00     | 0.09  | -0.05 | NA            |
| NO        | -0.07    | 0.21  | -0.07 | 0.02          |
| HNO       | -0.04    | 0.24  | -0.04 | 0.01          |
| NO$_2$    | 0.00     | -0.09 | 0.08  | -0.04         |
| NO$_3$    | -0.19    | -0.26 | -0.07 | 0.02          |
| trans-HNO$_2$ | 0.02 | 0.00  | 0.00  | -0.02         |
| cis-HNO$_2$  | 0.02 | 0.00  | 0.00  | NA            |
| HNO$_3$   | 0.00     | 0.00  | 0.02  | 0.01          |
| N$_2$O    | 0.00     | 0.06  | -0.07 | -0.09         |
| cis-N$_2$O$_2$ | -0.10 | 0.00  | -0.06 | 0.12          |
| N$_2$O$_3$ | 0.06    | 0.04  | 0.07  | 0.06          |
| N$_2$O$_4$ | 0.02    | -0.14 | 0.00  | -0.13         |
S6. Enthalpic effects

A potential source of discrepancies between the experimental and DFT-calculated enthalpies of formation might be the difference in temperature, as experimental measurements are performed at 298.15 K and DFT calculations are computed at 0 K. In case these contributions are included, Equation 3 in the main text would be modified by adding the change with temperature of the formation enthalpy:

$$
\Delta_f G^{DFT} \approx \Delta_f E^{DFT} + \Delta_f ZPE + (\Delta_f H_{@298K} - \Delta_f H_{@0K}) - T \Delta_f S
$$

(S4)

For the studied dataset, the enthalpy contribution due to the change in temperature corresponds on average to -0.01 eV/atom, as shown in Table S9. The changes for N₂, H₂ and O₂ are zero at all temperatures.

| species      | \(\Delta_f H_{@298K} - \Delta_f H_{@0K} / eV\) | \(\frac{1}{n}(\Delta_f H_{@298K} - \Delta_f H_{@0K}) / eV \text{ atom}^{-1}\) |
|--------------|---------------------------------------------|--------------------------------------------------------------------------------|
| NH₃          | -0.07                                       | -0.02                                                                         |
| NH₂OH        | -0.11                                       | -0.02                                                                         |
| NO           | 0.01                                        | 0.00                                                                          |
| HNO          | -0.03                                       | -0.01                                                                         |
| NO₂          | -0.03                                       | -0.01                                                                         |
| NO₃          | -0.05                                       | -0.01                                                                         |
| trans-HNO₂   | -0.06                                       | -0.02                                                                         |
| cis-HNO₂     | -0.07                                       | -0.02                                                                         |
| HNO₃         | -0.10                                       | -0.02                                                                         |
| N₂O          | -0.04                                       | -0.01                                                                         |
| cis-N₂O₂     | -0.02                                       | 0.00                                                                          |
| N₂O₃         | -0.05                                       | -0.01                                                                         |
| N₂O₄         | -0.10                                       | -0.02                                                                         |
| N₂O₅         | -0.10                                       | -0.01                                                                         |

Most enthalpic temperature corrections are smaller than -0.10 eV and there is a nearly perfect cancellation of these contributions when considering chemical reactions,
such that the change in the Gibbs energies due to temperature changes is negligible. For instance, consider the reaction: \( \text{NO}_2 + \text{NO}_3 \rightarrow \text{N}_2\text{O}_5 \). The change of enthalpy of this reaction from 0 to 298.15 K amounts to \(-0.10 - (-0.03) - (-0.05) = -0.02 \text{ eV}\). While in general we expect errors within \( \pm 0.05 \text{ eV} \) for the molecules in Table S9, the largest expectable error is 0.11 eV, which is the span of the values in Table S9 (i.e., NO and \( \text{NH}_2\text{OH} \)).

In conclusion, our method is based on free energies of formation, but its corrections are to be used for obtaining accurate reaction energies. Discarding enthalpic contributions due to the changes in temperature would typically lead to errors of magnitudes comparable to the MAEs and MAXs of the methods discussed in the main text. Thus, incorporating enthalpic corrections seems facultative at this point.

**S7. Analysis of the initial trends in Figure 5**

Equation 6 in the main text is:

\[
\Delta_f G_{\text{corrected}}^{\text{H}_2\text{N}_2\text{O}_2} = \Delta_f G_{\text{DFT}}^{\text{H}_2\text{N}_2\text{O}_2} + \frac{y}{2} \varepsilon_{\text{N}_2} + \frac{z}{2} \varepsilon_{\text{O}_2} - \varepsilon_{\text{H}_2\text{N}_2\text{O}_2}
\] (S5)

Before the errors in the compounds are detected, \( \varepsilon_{\text{H}_2\text{N}_2\text{O}_2} = 0 \). Besides, for PBE \( \varepsilon_{\text{N}_2} = 0.34 \text{ eV} \) and \( \varepsilon_{\text{O}_2} = -0.46 \text{ eV} \). The compounds with the smallest and largest number of atoms are NO and \( \text{N}_2\text{O}_5 \), respectively. For those, Equation S5 turns into:

\[
\Delta_f G_{\text{NO}}^{\text{corrected}} = \Delta_f G_{\text{DFT}}^{\text{NO}} + \frac{1\cdot0.34}{2} + \frac{1\cdot(-0.46)}{2} = \Delta_f G_{\text{DFT}}^{\text{NO}} - 0.06 \text{ eV}
\] (S6)

\[
\Delta_f G_{\text{N}_2\text{O}_5}^{\text{corrected}} = \Delta_f G_{\text{DFT}}^{\text{N}_2\text{O}_5} + \frac{2\cdot0.34}{2} + \frac{5\cdot(-0.46)}{2} = \Delta_f G_{\text{DFT}}^{\text{N}_2\text{O}_5} - 0.81 \text{ eV}
\] (S7)

Because the offsets are in the range of -0.06 to -0.81 eV, it is likely that upon correcting \( \text{O}_2 \) and \( \text{N}_2 \) the compounds will be below the parity line in Figure 5.

**S8. Direct coordinates**

| N  | E  | H  | 1 | 1 | 3 |
|----|----|----|---|---|---|
| NO | 0.3163211844490530 | 0.20564901711342 | 0.1366655811667420 | T | T | T |
| 0.266728304022136 | 0.29937019524365 | 0.134804084871437 | T | T | T |
| 0.16470121493093 | 0.188367649709510 | 0.216463313465539 | T | T | T |

**NH2OH**

| 1.00000000000000 | 0.00000000000000 | 0.00000000000000 | 0.00000000000000 | 0.00000000000000 | 0.00000000000000 |
| 0.2545210566049662 | 0.31994151585952 | 0.1363054226666130 | T | T | T |
| 0.164593323953584 | 0.2166596506103559 | 0.22641907963037 | T | T | T |

**Selective dynamics**

Direct

| 0.113333333333333 | 0.00000000000000 | 0.00000000000000 | 0.00000000000000 | 0.00000000000000 | 0.00000000000000 |
| 0.2545210566049662 | 0.31994151585952 | 0.1363054226666130 | T | T | T |
| 0.164593323953584 | 0.2166596506103559 | 0.22641907963037 | T | T | T |

N O H
References

(1) Haynes, W. M.; Lide, D. R.; Bruno, T. J. CRC Handbook of Chemistry and Physics, 97th ed.; CRC Press/Taylor And Francis: Boca Raton, FL, 2016. https://doi.org/10.1201/9781315380476.

(2) Linstrom, P. J.; Mallard, W. G. NIST Chemistry WebBook, NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 20899, 2022. https://doi.org/10.18434/T4D303.
(3) GAMS Development Corporation. *GAMS Documentation 39.1.1.*
https://www.gams.com/latest/docs/ (accessed 2022-06-04).

(4) Urrego-Ortiz, R.; Builes, S.; Calle-Vallejo, F. Impact of Intrinsic Density Functional Theory Errors on the Predictive Power of Nitrogen Cycle Electrocatalysis Models. *ACS Catal.* 2022, 12 (8), 4784–4791. https://doi.org/10.1021/acscatal.1c05333.

(5) Kurth, S.; Perdew, J. P.; Blaha, P. Molecular and Solid-State Tests of Density Functional Approximations: LSD, GGAs, and Meta-GGAs. *Int. J. Quantum Chem.* 1999, 75 (4-5), 889–909. https://doi.org/10.1002/(SICI)1097-461X(1999)75:4/5<889::AID-QUA54>3.0.CO;2-8.

(6) Ruscic, B.; Bross, D. H. *Active Thermochemical Tables (ATcT) values based on ver. 1.122 of the Thermochemical Network (2016).* https://atct.anl.gov/ (accessed 2022-08-05).

(7) Ruscic, B.; Pinzon, R. E.; Laszewski, G. von; Kodeboyina, D.; Burcat, A.; Leahy, D.; Montoy, D.; Wagner, A. F. Active Thermochemical Tables: Thermochemistry for the 21st Century. *J. Phys. Conf. Ser.* 2005, 16, 561–570. https://doi.org/10.1088/1742-6596/16/1/078.