Merging of the 'Spline-Pointwise' and 'Morse/Long-Range' Potential Function Forms for Direct-Potential-Fit Data Analyses

Jason Tao,[a] Robert J. Le Roy[a] and Asen Pashov[b]

[a] Guelph-Waterloo Centre for Graduate Work in Chemistry and Biochemistry, University of Waterloo, Waterloo, Ontario, Canada

[b] Department of Physics, Sofia University, Sofia, Bulgaria

Research supported by the Natural Sciences and Engineering Research Council of Canada.

Objectives of Spectroscopic Data Analysis

1. To provide an accurate, compact, and comprehensive representation of experimental data.

2. To be able to interpolate reliably for missing observations within the data range.

3. To be able to provide realistic predictions in the 'extrapolation region' outside the range of existing data.

4. To provide reliable estimates of physically interesting molecular properties (e.g., bond lengths, force constants, intensities).

Objectives of Spectroscopic Data Analysis

1. To provide an accurate, compact, and comprehensive representation of experimental data.

2. To be able to interpolate reliably for missing observations within the data range.

3. To be able to provide realistic predictions in the 'extrapolation region' outside the range of existing data.

4. To provide reliable estimates of physically interesting molecular properties (e.g., bond lengths, force constants, intensities).

What is the best way of doing this?

Ans. By using a compact, analytical potential energy function

Objectives of Spectroscopic Data Analysis

1. To provide an accurate, compact, and comprehensive representation of experimental data.

2. To be able to interpolate reliably for missing observations within the data range.

3. To be able to provide realistic predictions in the 'extrapolation region' outside the range of existing data.

4. To provide reliable estimates of physically interesting molecular properties (e.g., bond lengths, force constants, intensities).

What is the best way of doing this?

Ans. By using a compact, analytical potential energy function

How do we determine this potential energy function?

Ans. By performing 'direct potential fits'

Direct Potential Fits

- Simulate level energies as eigenvalues of some parameterized analytic potential energy function $V(r;\{p_j\})$

- Partial derivatives of observables w.r.t. parameters p_j required for fitting are generated readily using the Hellmann-Feynman theorem:

$$\frac{\partial E(v,J)}{\partial p_j} = \left\langle \psi_{v,J} \left| \frac{\partial V(r;\{p_i\})}{\partial p_j} \right| \psi_{v,J} \right\rangle$$

- Compare predicted transition energies with experiment, and optimize potential parameters via an iterative least-squares fit

Direct Potential Fits

- Simulate level energies as eigenvalues of some parameterized analytic potential energy function $V(r;\{p_j\})$

- Partial derivatives of observables w.r.t. parameters p_j required for fitting are generated readily using the Hellmann-Feynman theorem:

$$\frac{\partial E(v,J)}{\partial p_j} = \left\langle \psi_{v,J} \middle| \frac{\partial V(r;\{p_i\})}{\partial p_j} \middle| \psi_{v,J} \right\rangle$$

- Compare predicted transition energies with experiment, and optimize potential parameters via an iterative least-squares fit

Challenge . . . to develop analytic potential function forms

∗ flexible enough to fully represent extensive high-resolution data

∗ robust and 'well behaved' (no spurious extrapolation behaviour)

∗ compact and portable – defined by a 'modest' no. of parameters

∗ incorporating appropriate physical limiting behaviour

Two successful approaches: 1. a **'spline-pointwise' potential**

2. a **global analytic function**

1. The Spline Pointwise Potential (SPP)

* $V(r)$ is represented by a cubic spline through a set of specified points
* The energies of the points are the fitted parameters
* Attach a long-range function at a chosen (*ad hoc*) radial distance $r_{\rm out}$
* Fits/adjusts long-range coefficients (and sometimes also $r_{\rm out}$) until a smooth connection is achieved.

1. The Spline Pointwise Potential (SPP)

* $V(r)$ is represented by a cubic spline through a set of specified points
* The energies of the points are the fitted parameters
* Attach a long-range function at a chosen (*ad hoc*) radial distance $r_{\rm out}$
* Fits/adjusts long-range coefficients (and sometimes also $r_{\rm out}$) until a smooth connection is achieved.

Advantages

- Smooth and very flexible function: readily able to fit to irregular potentials such as double-minimum or 'shelf' potentials
- Little interparameter correlation

1. The Spline Pointwise Potential (SPP)

* $V(r)$ is represented by a cubic spline through a set of specified points
* The energies of the points are the fitted parameters
* Attach a long-range function at a chosen (*ad hoc*) radial distance $r_{\rm out}$
* Fits/adjusts long-range coefficients (and sometimes also $r_{\rm out}$) until a smooth connection is achieved.

Advantages

- Smooth and very flexible function: readily able to fit to irregular potentials such as those with double minima or a 'shelf'
- Little interparameter correlation

Disadvantages

- Discontinuous derivatives at attachment to the extrapolation regions
- Third derivatives discontinuous at all spline points. Higher-order derivatives do not exist.
- Requires a large number of parameters/spline points ($\gtrsim 50$), each specified to many significant digits, making it inconvenient to copy and use

2. Global Analytic Morse/Long-Range (MLR) Function

$$V(r) = \mathfrak{D}_{\mathrm{e}} \left(1 - \frac{u_{\mathrm{LR}}(r)}{u_{\mathrm{LR}}(r_{\mathrm{e}})} \; e^{-\beta(r) \cdot y_p^{\mathrm{eq}}(r)} \right)^2$$

- $y_p^{\mathrm{eq}}(r)$ is the radial variable

$$y_p^{\mathrm{eq}}(r) = \frac{r^p - r_{\mathrm{e}}^p}{r^p + r_{\mathrm{e}}^p}$$

2. Global Analytic Morse/Long-Range (MLR) Function

$$V(r) = \mathfrak{D}_{\mathrm{e}} \left(1 - \frac{u_{\mathrm{LR}}(r)}{u_{\mathrm{LR}}(r_{\mathrm{e}})} \; e^{-\beta(r) \cdot y_p^{\mathrm{eq}}(r)} \right)^2$$

- $y_p^{\mathrm{eq}}(r)$ is the radial variable

$$y_p^{\mathrm{eq}}(r) = \frac{r^p - r_{\mathrm{e}}^p}{r^p + r_{\mathrm{e}}^p}$$

- $u_{\mathrm{LR}}(r)$ defines long-range behaviour

$$u_{\mathrm{LR}}(r) = \sum_{i=1}^{last} \frac{C_{m_i}}{r^{m_i}}$$

2. Global Analytic Morse/Long-Range (MLR) Function

$$V(r) = \mathfrak{D}_\mathrm{e}\left(1 - \frac{u_\mathrm{LR}(r)}{u_\mathrm{LR}(r_\mathrm{e})}\, e^{-\beta(r)\cdot y_p^\mathrm{eq}(r)}\right)^2$$

- $y_p^\mathrm{eq}(r)$ is the radial variable

$$y_p^\mathrm{eq}(r) = \frac{r^p - r_\mathrm{e}^p}{r^p + r_\mathrm{e}^p}$$

- $u_\mathrm{LR}(r)$ defines long-range behaviour

$$u_\mathrm{LR}(r) = \sum_{i=1}^{last} \frac{C_{m_i}}{r^{m_i}}$$

- $\beta(r)$ is the exponent coefficient function defined as

$$\beta(r) \;=\; \beta_p^q(r) \;=\; \beta_\infty\, y_p^\mathrm{ref}(r) \,+\, [1 - y_p^\mathrm{ref}(r)] \sum_{i=0}^{N} \beta_i\, [y_q^\mathrm{ref}(r)]^i$$

where the coefficients β_i are the fitting parameters and

$$\lim_{r\to\infty} \beta(r) \;=\; \beta_\infty \;\equiv\; \ln\left(\frac{2\mathfrak{D}_e}{u_\mathrm{LR}(r_\mathrm{e})}\right)$$

these definitions allows the long-range behaviour of the potential to be

$$V(r) \;\approx\; \mathfrak{D}_e - u_\mathrm{LR}(r)$$

2. Global Analytic Morse/Long-Range (MLR) Function

$$V(r) = \mathfrak{D}_{\rm e} \left(1 - \frac{u_{\rm LR}(r)}{u_{\rm LR}(r_{\rm e})} \; e^{-\beta(r)\cdot y_p^{\rm eq}(r)} \right)^2$$

- $y_p^{\rm eq}(r)$ is the radial variable

$$y_p^{\rm eq}(r) = \frac{r^p - r_{\rm e}^p}{r^p + r_{\rm e}^p} \qquad\qquad y_p^{\rm ref}(r) = \frac{r^p - r_{\rm ref}^p}{r^p + r_{\rm ref}^p}$$

- $u_{\rm LR}(r)$ defines long-range behaviour

$$u_{\rm LR}(r) = \sum_{i=1}^{last} \frac{C_{m_i}}{r^{m_i}}$$

- $\beta(r)$ is the exponent coefficient function defined as

$$\beta(r) \;=\; \beta_p^q(r) \;=\; \beta_\infty \, y_p^{\rm ref}(r) \;+\; [1 - y_p^{\rm ref}(r)] \sum_{i=0}^{N} \beta_i \, [y_q^{\rm ref}(r)]^i$$

where the coefficients β_i are the fitting parameters and

$$\lim_{r\to\infty} \beta(r) \;=\; \beta_\infty \;\equiv\; \ln\left(\frac{2\mathfrak{D}_e}{u_{\rm LR}(r_{\rm e})} \right)$$

these definitions allows the long-range behaviour of the potential to be

$$V(r) \;\approx\; \mathfrak{D}_e - u_{\rm LR}(r)$$

2. Global Analytic Morse/Long-Range (MLR) Function

$$V(r) = \mathfrak{D}_{\rm e} \left(1 - \frac{u_{\rm LR}(r)}{u_{\rm LR}(r_{\rm e})} \; e^{-\beta(r)\cdot y_p^{\rm eq}(r)} \right)^2$$

Advantages

- Incorporates physically meaningful quantities ($\mathfrak{D}_e$, $r_{\rm e}$, C_n) as fitting parameters in the algebraic form
- Function and all derivatives smooth everywhere
- Requires relatively few parameters a achieve a better fit to experimental data than other forms

Disadvantages

2. Global Analytic Morse/Long-Range (MLR) Function

$$V(r) = \mathfrak{D}_{\mathrm{e}} \left(1 - \frac{u_{\mathrm{LR}}(r)}{u_{\mathrm{LR}}(r_{\mathrm{e}})} \; e^{-\beta(r) \cdot y_p^{\mathrm{eq}}(r)} \right)^2$$

Advantages

- Incorporates physically meaningful quantities ($\mathfrak{D}_e$, r_{e}, C_n) as fitting parameters in the algebraic form
- Function and all derivatives smooth everywhere
- Requires relatively few parameters a achieve a better fit to experimental data than other forms

Disadvantages

- High correlation among parameters
- Difficulty accounting for abrupt changes in shape

Can we combine the advantages of both forms?

The Spline Exponent-MLR (SE-MLR)

Same structure as the MLR, except that it is $\beta(y_p^{\mathrm{ref}}(r))$ [rather than $V(r)$] which is defined as a spline function through a specified set of function values, and it can be written as:

$$\beta(r) \; = \; \sum_{k=1}^{n} S_k(r)\,\beta(r_k) \; = \; \sum_{k=1}^{n} S_k(r)\,\beta_k$$

- Combines the very high flexibility of an SPP-type potential with the seamless incorporation of theoretical long-range behaviour inherent in the MLR form.

The Spline Exponent-MLR (SE-MLR)

Same structure as the MLR, except that it is $\beta(y_p^{\rm ref}(r))$ [rather than $V(r)$] which is defined as a spline function through a specified set of function values, and it can be written as:

$$\beta(r) \;=\; \sum_{k=1}^{n} S_k(r)\,\beta(r_k) \;=\; \sum_{k=1}^{n} S_k(r)\,\beta_k$$

- Combines the very high flexibility of an SPP-type potential with the seamless incorporation of theoretical long-range behaviour inherent in the MLR form.

How do we use the SE-MLR?

1. Choose parameters defining $y_p^{\rm ref}(r)$ $\{p$ and $r_{\rm ref}\}$
2. Place the spline points $\{N,\, y_p^{\rm ref}(r_k)\,,\, \beta_k\}$
3. Fit to the data to optimize the β_k values
4. Consider the quality of fit (dimensionless root-mean-square deviation, $\overline{dd}$) and check the resulting potential for unphysical behaviour

Applications

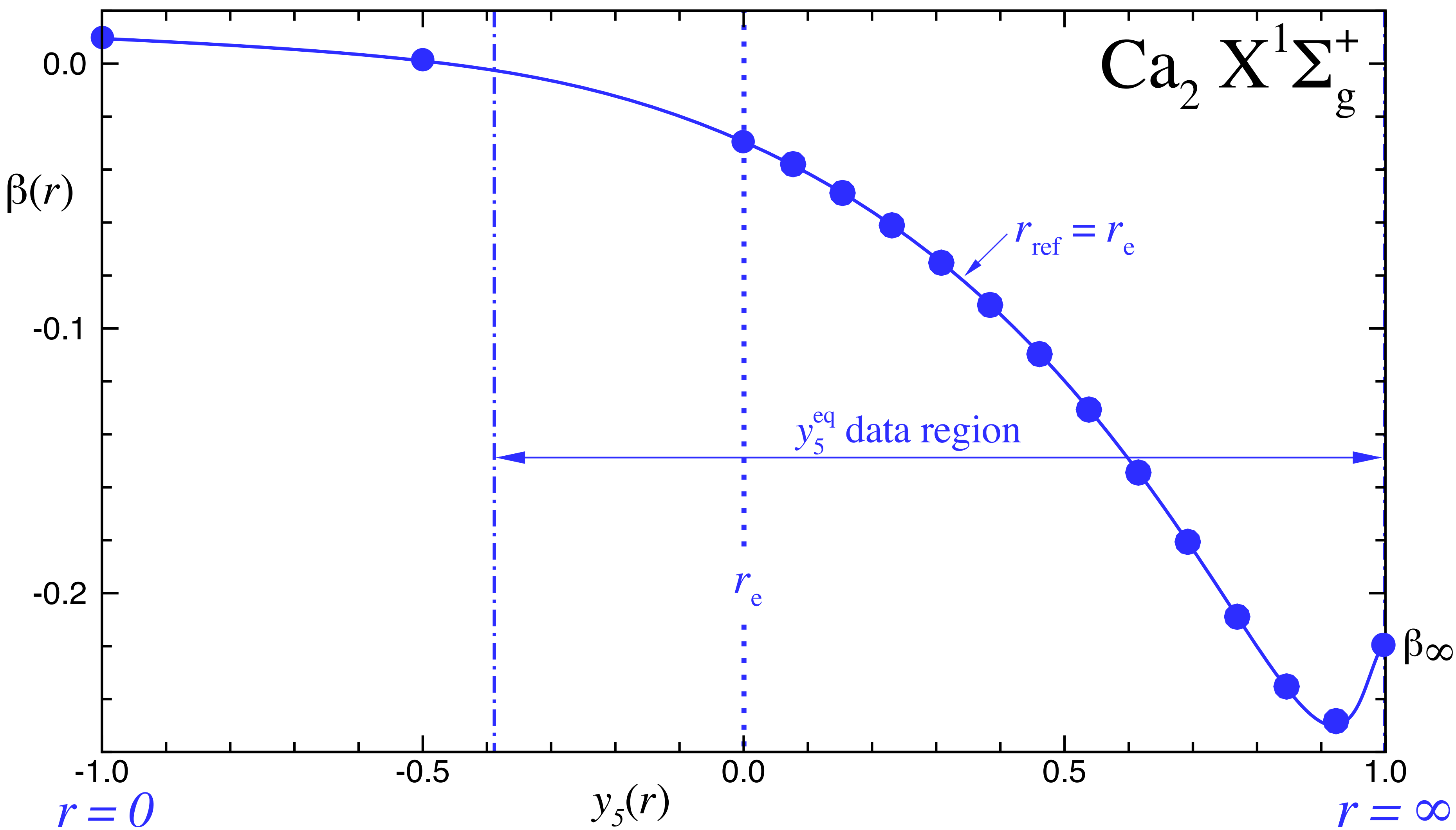
In order to test the abilities of the SE-MLR, consider the following systems:

1. $\mathbf{Ca}_2\ X^1\Sigma_g^+$
   - 3573 data, uncertainties 0.006-0.15 cm^{-1}
   - Data covers 99.97% of $\mathfrak{D}_e$ ($\sim$ 1100 cm^{-1})
   - Highest observed level ($v = 38$) bound by only $\sim 0.3\ \mathrm{cm}^{-1}$
   - MLR treatments fitted C_6 while holding other dispersion coefficients (C_8, C_{10}) fixed
2. $\mathbf{N}_2\ X^1\Sigma_g^+$
   - 1221 data, uncertainties 0.0015-0.015 cm^{-1}
   - Data covers only 47% of $\mathfrak{D}_e$
   - Highest observed level ($v = 20$) bound by 37600 cm^{-1}
   - ***Challenge*** Very narrow data region (0.9 - 1.55 Å), far extrapolation

How do we use the SE-MLR ?

1. Choose parameters defining $y_p^{\mathrm{ref}}(r)$ $\{\, p,\ r_{\mathrm{ref}}\}$

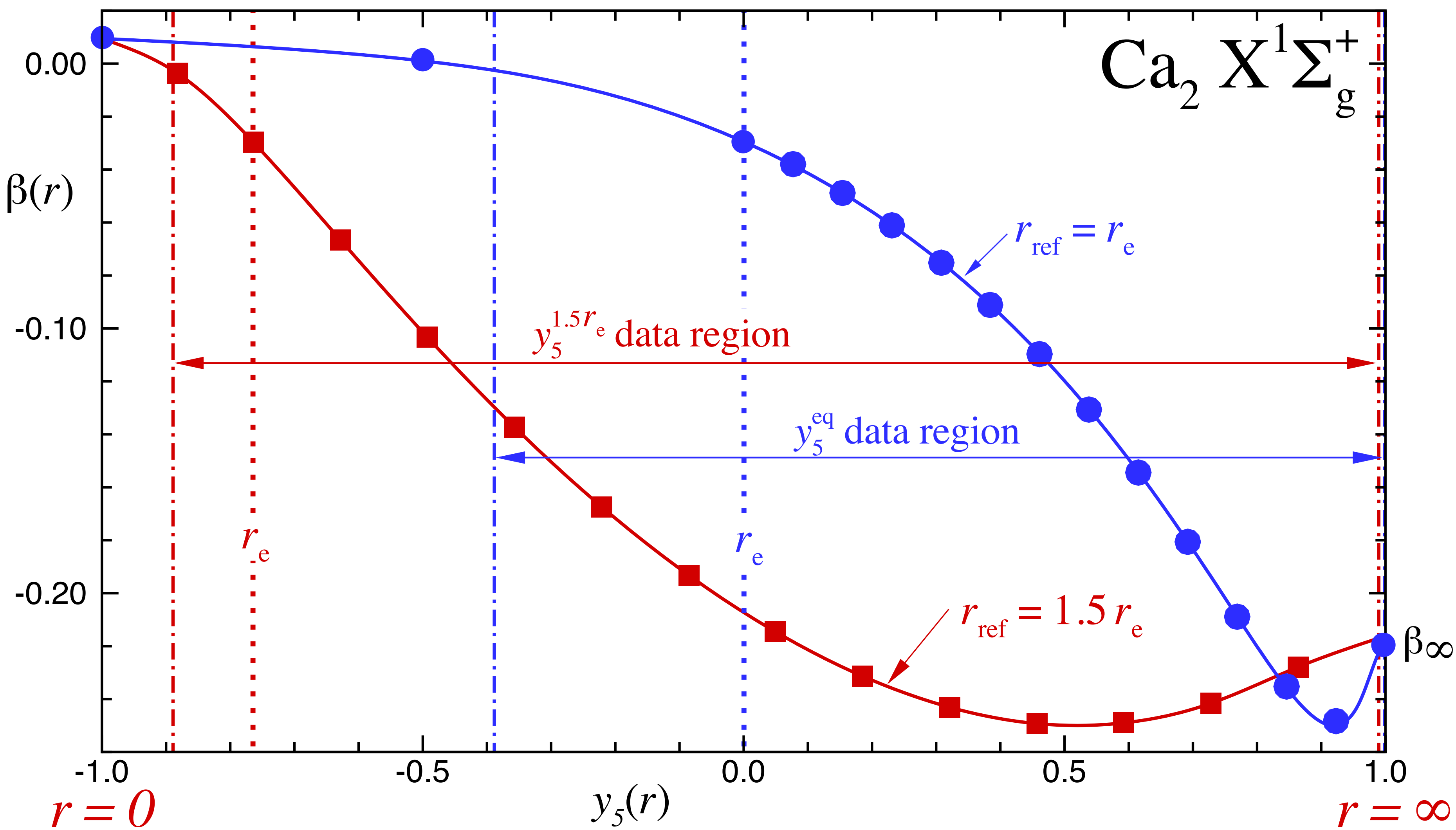
2. Place initial points $\{$here, 2 points $<\ r_{\mathrm{e}}$, 13 points $\geq r_{\mathrm{e}}$, $\beta(y_p^{\mathrm{ref}}(\infty)) = \beta_\infty\}$



How do we use the SE-MLR ?

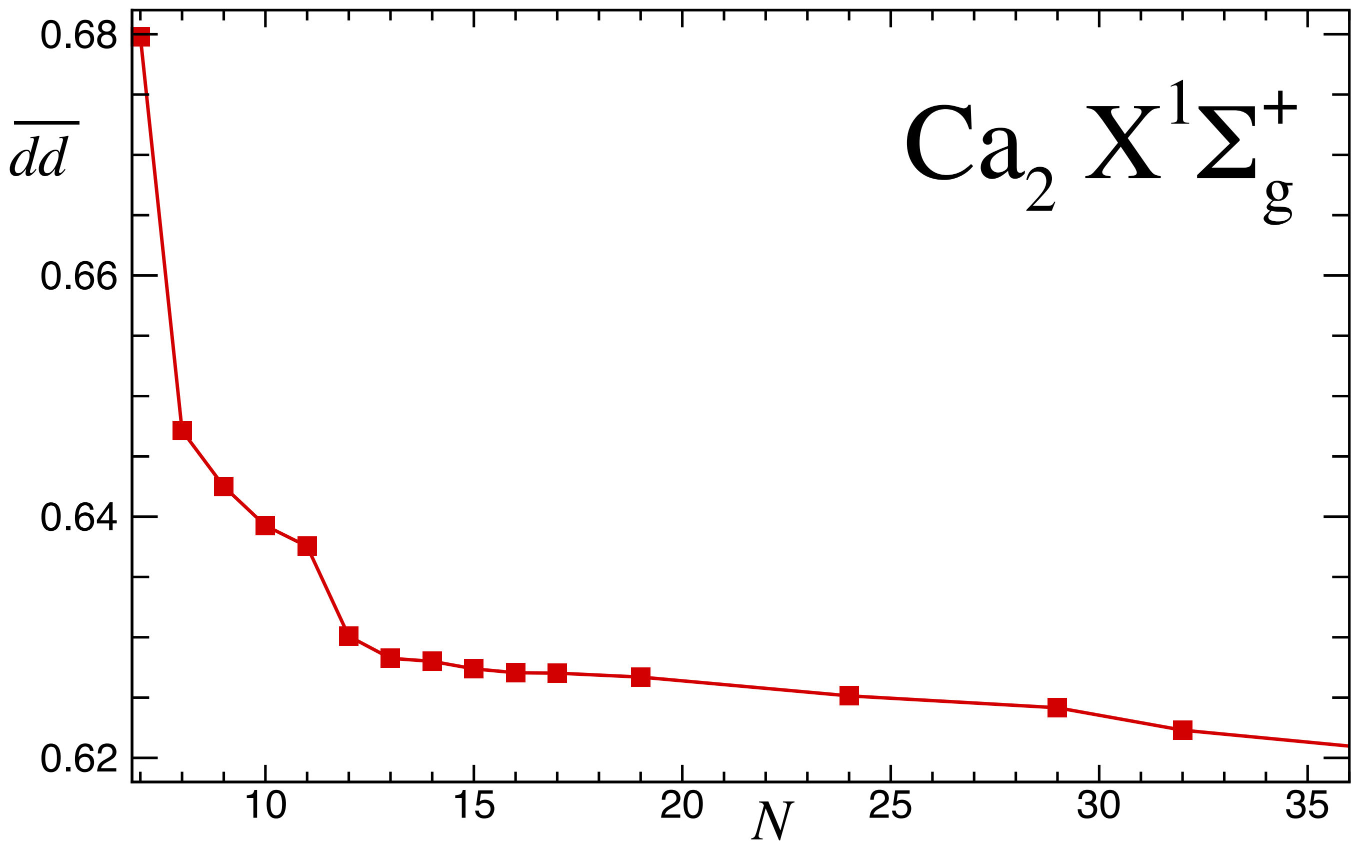
1. Choose parameters defining $y_p^{\mathrm{ref}}(r)$ $\{\ p,\ r_{\mathrm{ref}}\}$

2. Place initial points {here, 2 points $<\ r_{\mathrm{e}}$, 13 points $\geq r_{\mathrm{e}}$, $\beta(y_p^{\mathrm{ref}}(\infty)) = \beta_\infty$}



How do we use the SE-MLR?

1. Choose parameters defining $y_p^{\rm ref}(r)$ $\{p$ and $r_{\rm ref}\}$ √
2. Place the spline points $\{N,\ y_p^{\rm ref}(r_k)\ ,\ \beta_k\}$ √
3. Fit to the data to optimize the β_k values √
4. Consider the quality of fit (dimensionless root-mean-square deviation, $\overline{dd}$) and check the resulting potential for unphysical behaviour √
5. Change number of points and repeat
6. Compare models with different numbers of points



How do we use the SE-MLR?

1. Choose parameters defining $y_p^{\mathrm{ref}}(r)$ $\{p$ and $r_{\mathrm{ref}}\}$ $\surd$

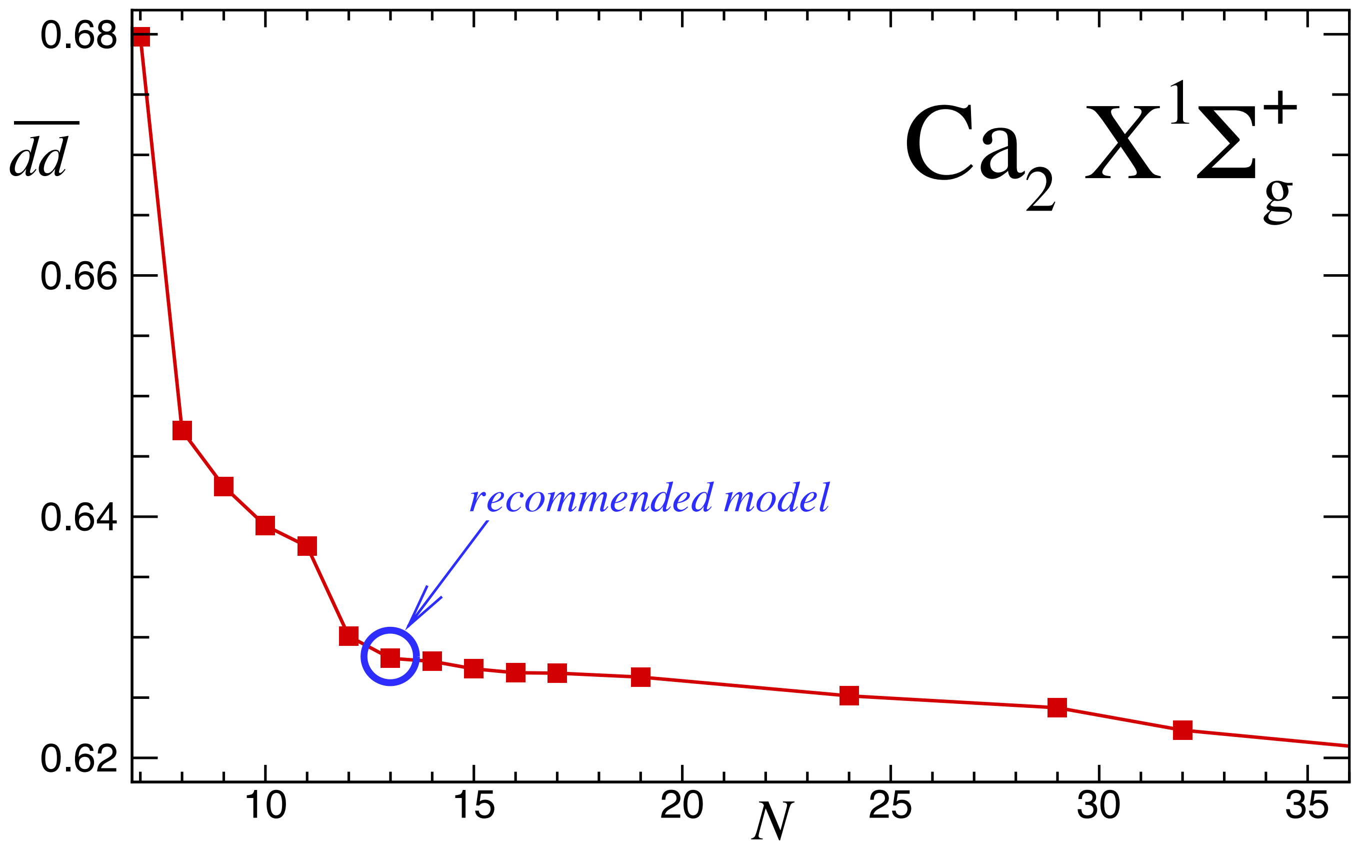
2. Place the spline points $\{N,\ y_p^{\mathrm{ref}}(r_k)\ ,\ \beta_k\}$ $\surd$

3. Fit to the data to optimize the β_k values $\surd$

4. Consider the quality of fit (dimensionless root-mean-square deviation, $\overline{dd}$) and check the resulting potential for unphysical behaviour $\surd$

5. Change number of points and repeat

6. Compare models with different numbers of points



Spline pointwise potential (2003)			
	$\mathfrak{D}_e$	1102.060	
	r_{out}	9.44	
	C_6	1.0023×10^7	
	C_8	3.808×10^8	
	C_{10}	$5.06\ \times10^9$	
r/Å	U/cm^{-1}	r/Å	U/cm^{-1}
3.096980	9246.6895	5.678571	636.3741
3.188725	6566.7325	5.809524	684.9589
3.280470	4525.7282	5.940476	728.9235
3.372215	3090.9557	6.071429	768.5976
3.463960	2134.2175	6.202381	804.2551
3.555705	1475.2425	6.333333	836.2419
3.647450	1004.5043	6.464286	864.8746
3.739195	661.4123	6.595238	890.4666
3.830940	410.6117	6.726191	913.2923
3.922685	234.0001	6.857143	933.6417
4.014430	116.0996	6.988095	951.7718
4.106174	44.5437	7.119048	967.8632
4.197920	8.6885	7.250000	982.2159
4.289664	0.1760	7.500000	1005.2497
4.381409	11.9571	7.750000	1023.6698
4.500000	48.5948	8.000000	1038.3262
4.630952	106.9081	8.358974	1054.3861
4.761905	175.7311	8.717949	1066.0579
4.892857	248.8199	9.076923	1074.5969
5.023809	322.3873	9.435897	1080.8961
5.154762	393.7222	9.794872	1085.5974
5.285714	461.4555	10.303419	1090.2990
5.416667	524.6311	10.811966	1093.5160
5.547619	582.9870	11.611111	1096.6870
$\overline{dd}$	0.70		

PE – MLR$_{5,3}$	
$\mathfrak{D}_e$	1102.081
r_e	4.27781
C_6	1.046×10^7
C_8	3.0608×10^8
C_{10}	8.344×10^9
r_{ref}	5.55
$\{p,q\}$	$\{5,3\}$
β_0	−0.19937072
β_1	−0.23219
β_2	−0.06091
β_3	0.1383
β_4	−0.1791
β_5	0.362
β_6	0.249
$\overline{dd}$	0.628

SE – MLR$_5$	
$\mathfrak{D}_e$	1102.072
r_e	4.27780
C_6	1.030×10^7
C_8	3.0608×10^8
C_{10}	8.344×10^9
r_{ref}	6.3
$y_5^{6.3}$	β
−1.000	0.0084239
−0.844	−0.0034414
−0.688	−0.0300237
−0.447	−0.0839978
−0.206	−0.1384960
0.034	−0.1865833
0.276	−0.2250030
0.517	−0.2481562
0.758	−0.2466246
1.000	−0.2134933
$\overline{dd}$	0.628

Results

The SE-MLR achieves accuracy of the PE-MLR with requiring significantly fewer parameters than the SPP, but more than the PE-MLR

$Ca_2(X^1\Sigma_g^+)$	SPP	PE-MLR	SE-MLR
# fitted param.	55	10	12
$\overline{dd}$	0.70	0.628	0.628

The fitted value of C_6 obtained using the SE-MLR shows less model dependence than when using the PE-MLR

Results

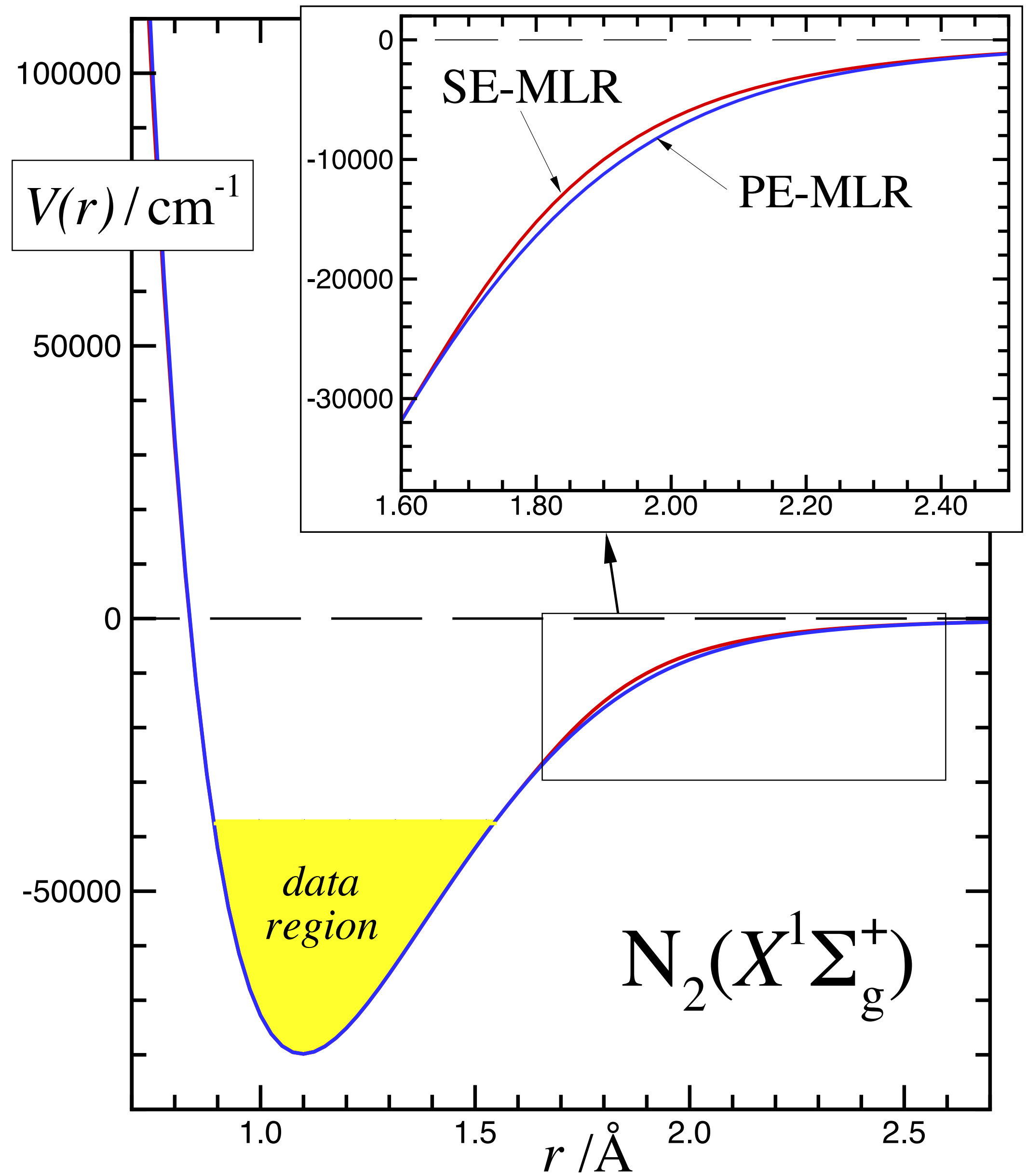
The SE-MLR achieves accuracy of the PE-MLR with requiring significantly fewer parameters than the SPP, but more than the PE-MLR

$Ca_2(X^1\Sigma_g^+)$	SPP	PE-MLR	SE-MLR
# fitted param.	55	10	12
$\overline{dd}$	0.70	0.628	0.628

The fitted value of C_6 obtained using the SE-MLR shows less model dependence than when using the PE-MLR

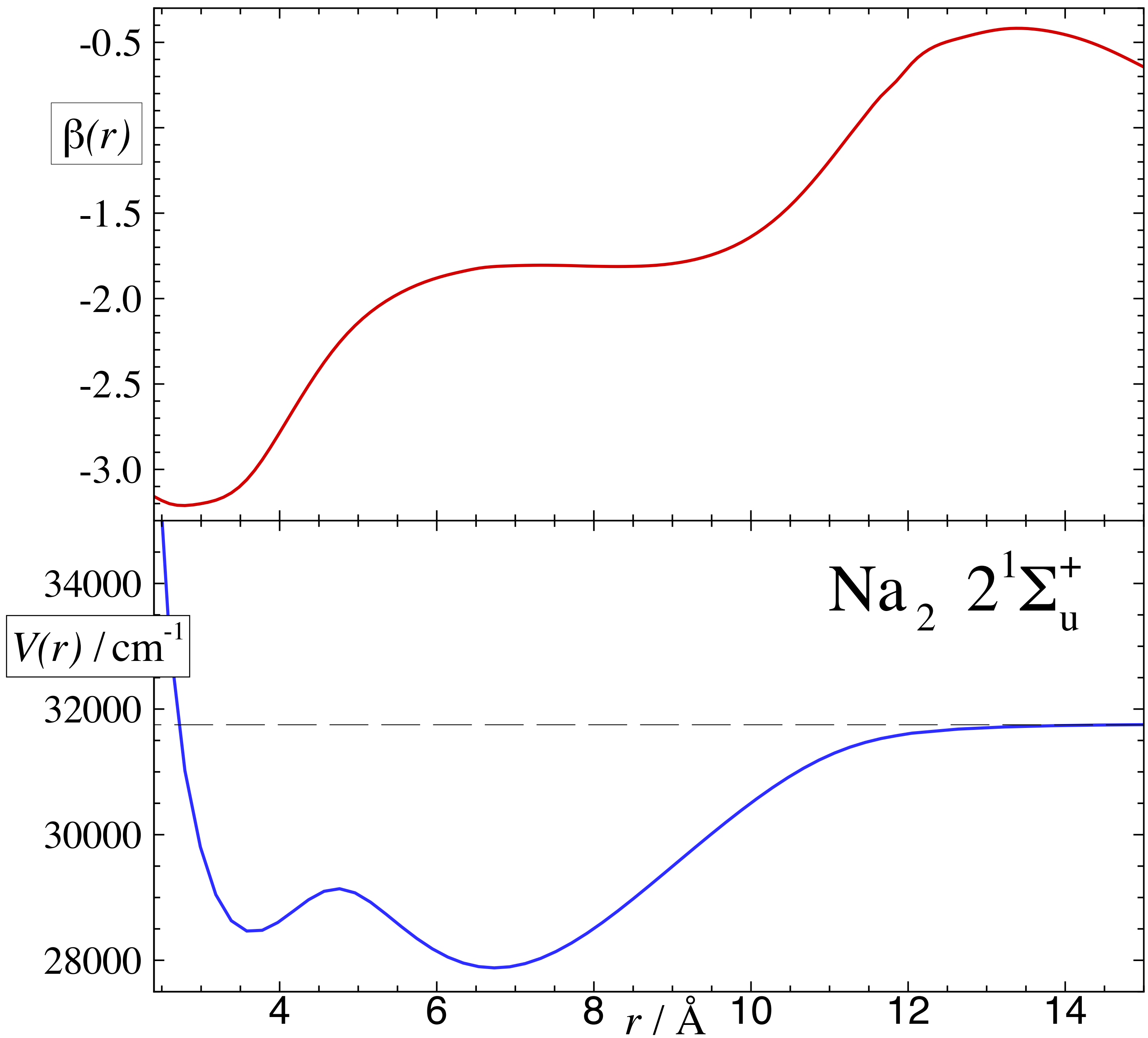
$N_2(X^1\Sigma_g^+$	PE-MLR	SE-MLR
# fitted param.	10	17
$\overline{dd}$	1.416	1.404

SE-MLR can incorporate a sensible long extrapolation to the limit, but a conventional SPP cannot !



A challenging Ssstem for the SE-MLR form

Double-minimum potential - $Na_2(2^1\Sigma_u^+)$



Conclusion

- By having the exponent coefficient function (rather than the potential itself) be represented by a cubic spline, the number of points required to describe the potential is reduced dramatically.

- The SE-MLR successfully combines the flexibility of the spline-pointwise approach with a natural incorporation of the theoretically predicted inverse-power long-range behaviour.

- To obtain a given quality of fit for a conventional single-minimum potential, the SE-MLR requires more parameters than does a PE-MLR.

- However, preliminary results suggest that fits using an SE-MLR may provide a more reliable determination of long-range coefficients such as C_6 .

Future Work

- Explore the SE-MLR's utility in describing double-minimum potentials.

- Possible quantitative incorporation of the correct, very short range 'united-atom-limit behaviour' $V(r) = \frac{Z_1 Z_2 e^2}{4\pi\epsilon_0 r}$?

Splitting the radial variable exponent

$$\beta(r) = c_0 y_q^{\rm ref}(r) + c_1 y_q^{\rm ref}(r) + c_2 y_q^{\rm ref}(r)^2 + c_3 y_q^{\rm ref}(r)^3$$

$$y_q^{\rm ref}(r) = \frac{1 - \left(\frac{r}{r_{\rm ref}}\right)^q}{1 + \left(\frac{r}{r_{\rm ref}}\right)^q} = \left(\frac{1}{1 + x^q}\right)(1 - x^q)$$

For $q < p$

$$e^{-\beta(y_q^{\rm ref}(r))y_p^{\rm eq}(r))} = e^{\beta_\infty}\left(1 + \frac{2r_{\rm ref}^q(c_1 + 2c_2 + 3c_3)}{r^q} + \frac{2r_{\rm eq}^p C}{r^p} + \frac{2r_{\rm ref}^{2q}(c_1 - 3c_3)}{r^{2q}}\right)$$

However, at $r = \infty$, $y_q^{\rm ref}(r) = 1$ and if

$$\frac{d\beta(y_q^{\rm ref})}{dy_q^{\rm ref}} = c_1 + 2c_2 + 3c_3 = 0$$

then at large distances

$$V(r) \simeq \mathfrak{D}_e\left(1 - u_{\rm LR(r)} + \frac{2r_{\rm e}^p\beta_\infty}{r^p} + \ldots\right)$$

Improving Short-range behaviour

$$\lim_{r\to 0}$$

$$\lim_{r\to 0} V(r) = \frac{C_1}{r} = \frac{Z_1 Z_2 C_1^{\mathrm{pp}}}{r} \approx \frac{Z_1 Z_2 C_1^{\mathrm{pp}} \mathfrak{D}_e u_{\mathrm{LR}}(r)^2 e^{2\beta(r=0)}}{u_{\mathrm{LR}}(r_{\mathrm{e}})^2 r}$$

$$\beta(r=0) = \ln\left(\frac{u_{\mathrm{LR}}\sqrt{Z_1 Z_2 C_1^{\mathrm{pp}}/\mathfrak{D}_e}\, u_{\mathrm{LR}}(r_{\mathrm{e}})}{\sum_{i=1}^{\mathrm{last}} \left(\frac{b\rho}{m_i}\right)^{m_i - 1/2}}\right)$$