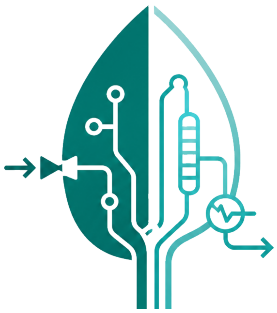

Choupo

The Open Source Chemical Process Simulator

Properties Guide

Version 2607

30 June 2026

Copyright © 2026 Vítor Geraldes, Pedro Mendes, and Miguel Rodrigues.

Authors: Vítor Geraldes, Pedro Mendes, and Miguel Rodrigues.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Code examples, Choupo case files, and other machine-readable examples are licensed under GPL-3.0-or-later.

Typeset in LaTeX.

Acknowledgment. Choupo was conceived and architected by Vítor Geraldes. Substantial parts of the initial implementation, documentation, and tutorial corpus were produced with assistance from Anthropic Claude Code using Claude Opus/Fable models. This guide is published under human curation, review, and editorial responsibility by Vítor Geraldes, Pedro Mendes, and Miguel Rodrigues.

License

Attribution-ShareAlike 4.0 International

===

Creative Commons Corporation ("Creative Commons") is not a law firm and
does not provide legal services or legal advice. Distribution of
Creative Commons public licenses does not create a lawyer-client or
other relationship. Creative Commons makes its licenses and related
information available on an "as-is" basis. Creative Commons gives no
warranties regarding its licenses, any material licensed under their
terms and conditions, or any related information. Creative Commons
disclaims all liability for damages resulting from their use to the
fullest extent possible.

Using Creative Commons Public Licenses

Creative Commons public licenses provide a standard set of terms and
conditions that creators and other rights holders may use to share
original works of authorship and other material subject to copyright
and certain other rights specified in the public license below. The
following considerations are for informational purposes only, are not
exhaustive, and do not form part of our licenses.

     Considerations for licensors: Our public licenses are
     intended for use by those authorized to give the public
     permission to use material in ways otherwise restricted by
     copyright and certain other rights. Our licenses are
     irrevocable. Licensors should read and understand the terms
     and conditions of the license they choose before applying it.
     Licensors should also secure all rights necessary before
     applying our licenses so that the public can reuse the
     material as expected. Licensors should clearly mark any
     material not subject to the license. This includes other CC-
     licensed material, or material used under an exception or
     limitation to copyright. More considerations for licensors:
    wiki.creativecommons.org/Considerations_for_licensors

     Considerations for the public: By using one of our public
     licenses, a licensor grants the public permission to use the
     licensed material under specified terms and conditions. If
     the licensor's permission is not necessary for any reason--for
     example, because of any applicable exception or limitation to
     copyright--then that use is not regulated by the license. Our
     licenses grant only permissions under copyright and certain
     other rights that a licensor has authority to grant. Use of
     the licensed material may still be restricted for other
     reasons, including because others have copyright or other
     rights in the material. A licensor may make special requests,
     such as asking that all changes be marked or described.
     Although not required by our licenses, you are encouraged to
     respect those requests where reasonable. More considerations
     for the public:
    wiki.creativecommons.org/Considerations_for_licensees

===

Creative Commons Attribution-ShareAlike 4.0 International Public
License

By exercising the Licensed Rights (defined below), You accept and agree
to be bound by the terms and conditions of this Creative Commons
Attribution-ShareAlike 4.0 International Public License ("Public
License"). To the extent this Public License may be interpreted as a
contract, You are granted the Licensed Rights in consideration of Your
acceptance of these terms and conditions, and the Licensor grants You
such rights in consideration of benefits the Licensor receives from
making the Licensed Material available under these terms and
conditions.

Section 1 -- Definitions.

  a. Adapted Material means material subject to Copyright and Similar
     Rights that is derived from or based upon the Licensed Material
     and in which the Licensed Material is translated, altered,
     arranged, transformed, or otherwise modified in a manner requiring
     permission under the Copyright and Similar Rights held by the
     Licensor. For purposes of this Public License, where the Licensed
     Material is a musical work, performance, or sound recording,
     Adapted Material is always produced where the Licensed Material is
     synched in timed relation with a moving image.

  b. Adapter's License means the license You apply to Your Copyright
     and Similar Rights in Your contributions to Adapted Material in
     accordance with the terms and conditions of this Public License.

  c. BY-SA Compatible License means a license listed at
     creativecommons.org/compatiblelicenses, approved by Creative
     Commons as essentially the equivalent of this Public License.

  d. Copyright and Similar Rights means copyright and/or similar rights
     closely related to copyright including, without limitation,
     performance, broadcast, sound recording, and Sui Generis Database
     Rights, without regard to how the rights are labeled or
     categorized. For purposes of this Public License, the rights
     specified in Section 2(b)(1)-(2) are not Copyright and Similar

Rights.

e. Effective Technological Measures means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.

f. Exceptions and Limitations means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.

g. License Elements means the license attributes listed in the name of a Creative Commons Public License. The License Elements of this Public License are Attribution and ShareAlike.

h. Licensed Material means the artistic or literary work, database, or other material to which the Licensor applied this Public License.

i. Licensed Rights means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.

j. Licensor means the individual(s) or entity(ies) granting rights under this Public License.

k. Share means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.

l. Sui Generis Database Rights means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

m. You means the individual or entity exercising the Licensed Rights under this Public License. Your has a corresponding meaning.

Section 2 -- Scope.

a. License grant.

1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:

a. reproduce and Share the Licensed Material, in whole or in part; and

b. produce, reproduce, and Share Adapted Material.

2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.

3. Term. The term of this Public License is specified in Section 6(a).

4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this Public License, simply making modifications authorized by this Section 2(a) (4) never produces Adapted Material.

5. Downstream recipients.

a. Offer from the Licensor -- Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

b. Additional offer from the Licensor -- Adapted Material. Every recipient of Adapted Material from You automatically receives an offer from the Licensor to exercise the Licensed Rights in the Adapted Material under the conditions of the Adapter's License You apply.

c. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

6. No endorsement. Nothing in this Public License constitutes or
   may be construed as permission to assert or imply that You
   are, or that Your use of the Licensed Material is, connected
   with, or sponsored, endorsed, or granted official status by,
   the Licensor or others designated to receive attribution as
   provided in Section 3(a)(1)(A)(i).

b. Other rights.

1. Moral rights, such as the right of integrity, are not
   licensed under this Public License, nor are publicity,
   privacy, and/or other similar personality rights; however, to
   the extent possible, the Licensor waives and/or agrees not to
   assert any such rights held by the Licensor to the limited
   extent necessary to allow You to exercise the Licensed
   Rights, but not otherwise.

2. Patent and trademark rights are not licensed under this
   Public License.

3. To the extent possible, the Licensor waives any right to
   collect royalties from You for the exercise of the Licensed
   Rights, whether directly or through a collecting society
   under any voluntary or waivable statutory or compulsory
   licensing scheme. In all other cases the Licensor expressly
   reserves any right to collect such royalties.

Section 3 -- License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the
following conditions.

a. Attribution.

1. If You Share the Licensed Material (including in modified
   form), You must:

   a. retain the following if it is supplied by the Licensor
      with the Licensed Material:

      i. identification of the creator(s) of the Licensed
         Material and any others designated to receive
         attribution, in any reasonable manner requested by
         the Licensor (including by pseudonym if
         designated);

      ii. a copyright notice;

      iii. a notice that refers to this Public License;

      iv. a notice that refers to the disclaimer of
          warranties;

      v. a URI or hyperlink to the Licensed Material to the
         extent reasonably practicable;

   b. indicate if You modified the Licensed Material and
      retain an indication of any previous modifications; and

   c. indicate the Licensed Material is licensed under this
      Public License, and include the text of, or the URI or
      hyperlink to, this Public License.

2. You may satisfy the conditions in Section 3(a)(1) in any
   reasonable manner based on the medium, means, and context in
   which You Share the Licensed Material. For example, it may be
   reasonable to satisfy the conditions by providing a URI or
   hyperlink to a resource that includes the required
   information.

3. If requested by the Licensor, You must remove any of the
   information required by Section 3(a)(1)(A) to the extent
   reasonably practicable.

b. ShareAlike.

   In addition to the conditions in Section 3(a), if You Share
   Adapted Material You produce, the following conditions also apply.

1. The Adapter's License You apply must be a Creative Commons
   license with the same License Elements, this version or
   later, or a BY-SA Compatible License.

2. You must include the text of, or the URI or hyperlink to, the
   Adapter's License You apply. You may satisfy this condition
   in any reasonable manner based on the medium, means, and
   context in which You Share Adapted Material.

3. You may not offer or impose any additional or different terms
   or conditions on, or apply any Effective Technological
   Measures to, Adapted Material that restrict exercise of the
   rights granted under the Adapter's License You apply.

Section 4 -- Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that
apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right
   to extract, reuse, reproduce, and Share all or a substantial
   portion of the contents of the database;

b. if You include all or a substantial portion of the database
   contents in a database in which You have Sui Generis Database
   Rights, then the database in which You have Sui Generis Database
   Rights (but not its individual contents) is Adapted Material,
   including for purposes of Section 3(b); and

c. You must comply with the conditions in Section 3(a) if You Share
   all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not
replace Your obligations under this Public License where the Licensed
Rights include other Copyright and Similar Rights.

Section 5 -- Disclaimer of Warranties and Limitation of Liability.

a. UNLESS OTHERWISE SEPARATELY UNDERTAKEN BY THE LICENSOR, TO THE
   EXTENT POSSIBLE, THE LICENSOR OFFERS THE LICENSED MATERIAL AS-IS
   AND AS-AVAILABLE, AND MAKES NO REPRESENTATIONS OR WARRANTIES OF
   ANY KIND CONCERNING THE LICENSED MATERIAL, WHETHER EXPRESS,
   IMPLIED, STATUTORY, OR OTHER. THIS INCLUDES, WITHOUT LIMITATION,
   WARRANTIES OF TITLE, MERCHANTABILITY, FITNESS FOR A PARTICULAR
   PURPOSE, NON-INFRINGEMENT, ABSENCE OF LATENT OR OTHER DEFECTS,
   ACCURACY, OR THE PRESENCE OR ABSENCE OF ERRORS, WHETHER OR NOT
   KNOWN OR DISCOVERABLE. WHERE DISCLAIMERS OF WARRANTIES ARE NOT
   ALLOWED IN FULL OR IN PART, THIS DISCLAIMER MAY NOT APPLY TO YOU.

b. TO THE EXTENT POSSIBLE, IN NO EVENT WILL THE LICENSOR BE LIABLE
   TO YOU ON ANY LEGAL THEORY (INCLUDING, WITHOUT LIMITATION,
   NEGLIGENCE) OR OTHERWISE FOR ANY DIRECT, SPECIAL, INDIRECT,
   INCIDENTAL, CONSEQUENTIAL, PUNITIVE, EXEMPLARY, OR OTHER LOSSES,
   COSTS, EXPENSES, OR DAMAGES ARISING OUT OF THIS PUBLIC LICENSE OR
   USE OF THE LICENSED MATERIAL, EVEN IF THE LICENSOR HAS BEEN
   ADVISED OF THE POSSIBILITY OF SUCH LOSSES, COSTS, EXPENSES, OR
   DAMAGES. WHERE A LIMITATION OF LIABILITY IS NOT ALLOWED IN FULL OR
   IN PART, THIS LIMITATION MAY NOT APPLY TO YOU.

c. The disclaimer of warranties and limitation of liability provided
   above shall be interpreted in a manner that, to the extent
   possible, most closely approximates an absolute disclaimer and
   waiver of all liability.

Section 6 -- Term and Termination.

a. This Public License applies for the term of the Copyright and
   Similar Rights licensed here. However, if You fail to comply with
   this Public License, then Your rights under this Public License
   terminate automatically.

b. Where Your right to use the Licensed Material has terminated under
   Section 6(a), it reinstates:

   1. automatically as of the date the violation is cured, provided
      it is cured within 30 days of Your discovery of the
      violation; or

   2. upon express reinstatement by the Licensor.

   For the avoidance of doubt, this Section 6(b) does not affect any
   right the Licensor may have to seek remedies for Your violations
   of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the
   Licensed Material under separate terms or conditions or stop
   distributing the Licensed Material at any time; however, doing so
   will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public
   License.

Section 7 -- Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different
   terms or conditions communicated by You unless expressly agreed.

b. Any arrangements, understandings, or agreements regarding the
   Licensed Material not stated herein are separate from and
   independent of the terms and conditions of this Public License.

Section 8 -- Interpretation.

a. For the avoidance of doubt, this Public License does not, and
   shall not be interpreted to, reduce, limit, restrict, or impose
   conditions on any use of the Licensed Material that could lawfully
   be made without permission under this Public License.

b. To the extent possible, if any provision of this Public License is
   deemed unenforceable, it shall be automatically reformed to the
   minimum extent necessary to make it enforceable. If the provision
   cannot be reformed, it shall be severed from this Public License
   without affecting the enforceability of the remaining terms and
   conditions.

c. No term or condition of this Public License will be waived and no
   failure to comply consented to unless expressly agreed to by the
   Licensor.

d. Nothing in this Public License constitutes or may be interpreted
   as a limitation upon, or waiver of, any privileges and immunities
   that apply to the Licensor or You, including from the legal
   processes of any jurisdiction or authority.

===

Creative Commons is not a party to its public
licenses. Notwithstanding, Creative Commons may elect to apply one of
its public licenses to material it publishes and in those instances
will be considered the "Licensor." The text of the Creative Commons
public licenses is dedicated to the public domain under the CC0 Public
Domain Dedication. Except for the limited purpose of indicating that
material is shared under a Creative Commons public license or as
otherwise permitted by the Creative Commons policies published at
creativecommons.org/policies, Creative Commons does not authorize the
use of the trademark "Creative Commons" or any other trademark or logo
of Creative Commons without its prior written consent including,
without limitation, in connection with any unauthorized modifications
to any of its public licenses or any other arrangements,
understandings, or agreements concerning use of licensed material. For
the avoidance of doubt, this paragraph does not form part of the
public licenses.

Creative Commons may be contacted at creativecommons.org.

Trademarks

Choupo is a trademark of TalentGround Lda.

Anthropic, Claude, and Claude Code are trademarks or registered trademarks of Anthropic, PBC.

NIST WebBook is a product or service of the National Institute of Standards and Technology.

Contents

"Thermodynamics is a funny subject. The first time you go through it, you don't understand it at all. The second time you go through it, you think you understand it, except for one or two points. The third time you go through it, you know you don't understand it, but by that time you are so used to it, it doesn't bother you any more."

— Arnold Sommerfeld

1 Introduction: the difficulty, the map, the promise

What this guide is — and is not. This is the props-**mode workflow** guide: how you *estimate* a missing compound, *fit* its binary pairs, and *consistency-test* the data — the see-then-decide loop of the Props/Explore workspace. It is *not* the property *theory*. For the first-principles derivations of NRTL, Wilson, UNIQUAC, UNIFAC, the cubic equations of state (SRK, Peng–Robinson), the electrolyte models (Pitzer, eNRTL, aqueous speciation), Henry's law, Rackett, Ambrose–Walton, sublimation and the NASA–CEA Gibbs-of-formation data, see the **Theory Guide** (the in-GUI model panels deep-link straight into it).

If thermodynamics felt slippery the first time — the formalism pristine, yet somehow never self-sufficient when you sat down with real data — you are in good company: Sommerfeld, who reshaped quantum physics, said he never wrote his thermodynamics book because he never felt he had understood the subject. The difficulty is real, but it is not formal incompleteness. It is that an *axiomatic-looking* macroscopic theory rests, underneath, on emergent statistical facts and on parameters that must be *measured*, not derived. The confusion is the seam between the two showing through. This guide does not hide that seam — it makes it the organising idea.

1.1 The target: where theory grips

Picture property prediction as a target (Figure 1): pristine at the core, increasingly empirical toward the rim.

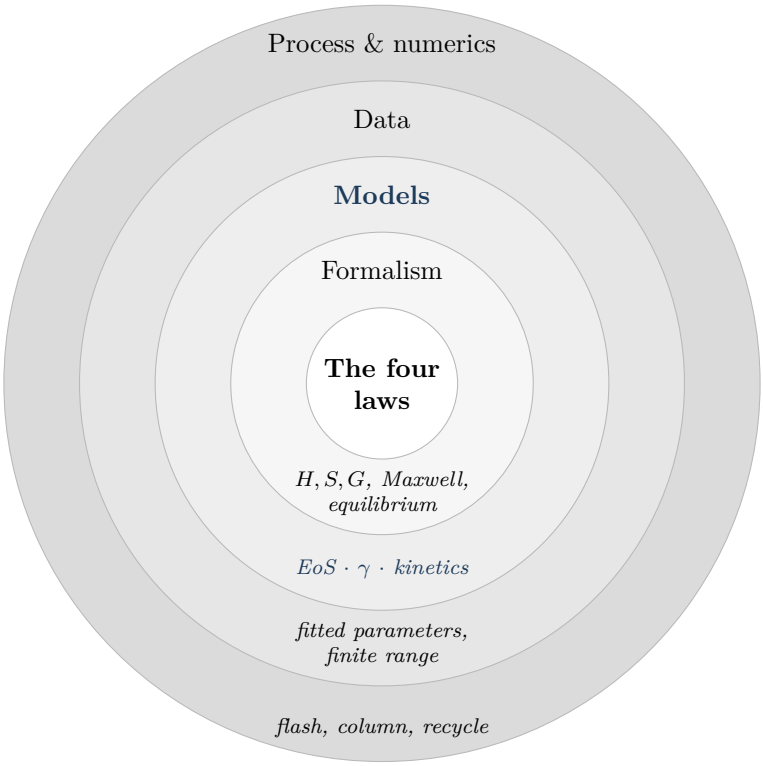

Figure 1: The layers of property prediction. The two inner shells follow from the laws alone. The accent shell — *constitutive models* — is the first fault line: the laws say *that* fugacities must match across phases, never *how* to compute one for a real mixture. From there outward, the theory is necessary but no longer sufficient.

- **The four laws (core).** The zeroth, first, second and third laws define temperature, internal energy, entropy, and — via Nernst — an absolute zero of entropy. Parameter-free.
- **Formalism.** Pure manipulation yields the potentials H, G, A, the Maxwell relations, the equilibrium criteria (equality of fugacity across phases, $\Delta G = -RT \ln K$ for reaction) and the phase rule. Still derivable from nothing but the laws.
- **Constitutive models — where it first grips.** The laws require fugacities to be equal across phases; they do not tell you how to *compute* a fugacity for a real mixture. You must supply a model they do not give you — an equation of state, an activity model, a kinetic law. In Choupo: `idealGas/SRK/PR`; `ideal/NRTL/Wilson/UNIFAC`.
- **Data — "go to the laboratory."** Those models carry parameters thermodynamics cannot predict: T_c, P_c, ω, the Antoine constants, $C_p^\circ(T)$, ΔH_f°, S°, the NRTL a_{ij}. Each is measured or estimated, and each has a finite

validity range. In Choupo, these are the `.dat` files.

- **Process & numerics.** Where the predicted numbers are consumed — flash, distillation, recycle — and grip a second time, now numerically (multiple EoS roots, phase stability, initialisation).

Key idea. Sommerfeld's difficulty is this picture: a theory that *looks* axiomatic to the rim, but stops being self-sufficient at the accent shell. The honest move is not to paper over the seam between rigorous derivation and fitted correlation — it is to show exactly where it runs.

1.2 Where the models came from — a genealogy

The chain in §3 shows there are exactly *two* ways to a fugacity in a mixture: through an equation of state (φ) or through an activity coefficient (γ). Each has its own lineage, and knowing them explains why the model list looks the way it does — and why it is a *list* at all rather than a single answer.

The equation-of-state line (1873 →). Van der Waals wrote one equation for both phases, continuous through the critical point: the first time liquid and gas were the same substance in one formula. Redlich–Kwong (1949) improved the attraction term; Soave (1972) made it obey the vapour-pressure curve by introducing the acentric factor; Peng–Robinson (1976) repaired the liquid density. All four share one skeleton — repulsion plus attraction, with the parameters generated from T_c, P_c, ω — which is why adding a new cubic to Choupo touches *no data file*. Then the molecular branch: Wertheim's association theory (1984) led to SAFT (1989) and PC-SAFT (2001), where a molecule is a *chain of segments* and the parameters are $m, \sigma, \varepsilon/k$ instead of the critical triad. That is exactly why PC-SAFT extrapolates in temperature and density where a cubic drifts — and why it cannot borrow the cubic's parameters.

The activity-coefficient line (1964 →). Regular-solution theory failed on polar mixtures. Wilson (1964) introduced *local composition* — the admission that a molecule does not see its neighbours in bulk proportions — but could not describe a liquid–liquid split. NRTL (Renon & Prausnitz, 1968) added non-randomness and *could*: that is why an LLE case in this guide reaches for NRTL and not Wilson. UNIQUAC (Abrams & Prausnitz, 1975) split the excess Gibbs energy into a combinatorial part (size and shape) and a residual part (energy). UNIFAC (Fredenslund, 1975) made UNIQUAC *predictive* by decomposing molecules into functional groups, so a pair never measured could still be estimated. COSMO-RS (Klamt, 1995) and COSMO-SAC (Lin & Sandler, 2002) went further and replaced the group table with the molecule's own screening-charge surface from a quantum calculation — the first genuinely *a priori* route to γ, and the one that needs no binary datum at all.

The electrolyte line (1923 →). Debye–Hückel derived the ionic atmosphere and is *exact* in the dilute limit — and only there. Davies (1962) stretched it empirically to a few tenths molal. Pitzer (1973) abandoned the pretence of derivation and wrote a virial expansion in molality that holds to seawater and beyond. eNRTL (Chen, 1982) carried local composition into ionic solutions, which is what makes a mixed solvent — salt plus water plus ethanol — tractable at all.

Key idea. Each model was invented because the previous one *failed* at something specific. The list is not a menu of equivalent options; it is a history of failures, each one repaired. Choosing a property method is therefore choosing *which failure you can live with* in your system — which is why this architecture insists that the choice be declared in the case file, verified at assembly, and confessed in the log. A simulator that hides the choice behind a dropdown has not removed the failure; it has removed your view of it.

1.3 The glass-box promise

Because the outer shells are empirical, every number Choupo predicts should travel with two things attached:

- **Provenance** — which method produced it (rigorous derivation, fitted correlation, or group-contribution estimate) and the source of each parameter (a measured datum or an estimate).
- **Validity awareness** — whether a correlation is *extrapolating beyond its fitted range*, with a warning when it is, rather than a confident number where none is warranted.

Intuition. Ask Choupo for the vapour pressure of nitrogen at room temperature and it *refuses*: N_2 is supercritical there (its critical temperature is $-147\,°C$), and a saturation pressure does not exist above the critical point. The P^{sat} curve simply *breaks* at T_c instead of extrapolating Antoine into a meaningless $\sim$600 bar. That refusal is the whole promise in miniature: the model declines to invent a number the physics forbids.

1.4 Where we are heading

A single property over a wide temperature span is not one problem — it is several physical regimes stitched together, and your confidence in any predicted number rises and collapses as you cross them. Figure 2 is the destination of this guide: the vapour pressure of a **water / ethanol / benzene** mixture from −200 to 1000 °C, drawn not as a curve but as a band of *confidence*.

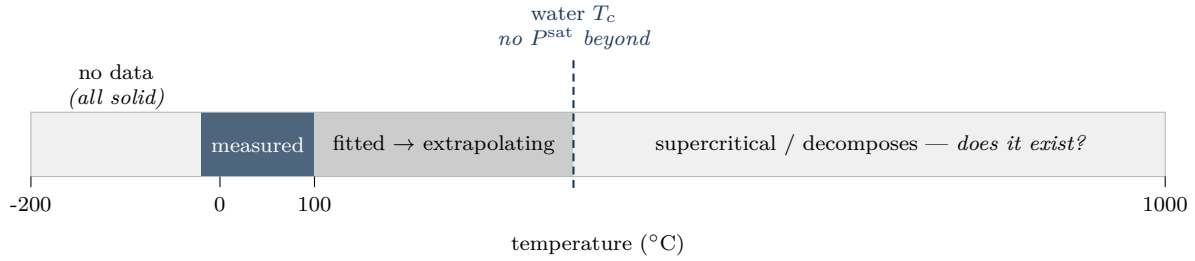

Figure 2: The same property, five honesties. A narrow window where we actually have data; degradation outward as the correlation extrapolates; a hard *wall* at the critical temperature where the property ceases to exist (the nitrogen refusal, generalised); and a cryogenic desert where every species has frozen solid (ethanol last, at −114 °C) and there is simply no data. The theory runs to both edges; reality and data do not.

We build each piece of that picture in the chapters that follow — estimating a missing component (§6), curating the mixture models against lab points (§7), and testing whether the data itself is even consistent (§8). We read that span regime by regime at the end of this introduction (§1.7), and return to solve it numerically, end to end, at the close.

1.5 Anatomy of a component: where the numbers come from

Open any `.dat` and you see a flat list of numbers. The natural question — the one this whole guide turns on — is: *which of these are independent inputs, and which derived properties does the engine actually build from them?* Let us read `acetone.dat` field by field and follow each number to the property it feeds and to the line of code that consumes it.

Field	What it is, and what it builds
Identity	
`MW`	molar mass; converts molar ↔ mass everywhere.
Caloric anchors — `standardThermochemistry`	
`dHf_298` ΔH_f°	the *position* of the enthalpy scale: bond energy C_p knows nothing about (combustion calorimetry).
`s_298` S°	the *absolute* (third-law) entropy at 298 K — a true zero-referenced value, not a difference.
Caloric shape — `idealGasHeatCapacity`	
`coefficients` $C_p^\circ(T)$	the master correlation: integrated to give the *temperature shape* of H, S, G (below).
Corresponding-states trio	
`Tc`, `Pc`	critical point: set the *size* of the cubic-EoS attraction term a_c, and the reduced temperature $T_r = T/T_c$.
`omega` ω	acentric factor: sets the *temperature shape* of that attraction, $\alpha(T_r;\omega)$ — how non-spherical / polar the molecule is.
Phase change	
`Tb`, `HvapTb`	boiling point + latent heat at T_b; with `Tc` drive Watson's $\Delta H_{vap}(T)$.
`Vliq`	liquid molar volume (Poynting, liquid density).
Correlations & recipes	
`vaporPressure`	Antoine $\log_{10} P^{sat} = A - B/(T+C)$, with its `Trange`; the saturation curve.
`groups`	the molecular decomposition a group-contribution estimator / UNIFAC reads (the curation recipe, §6).
`diffusionVolume`	Fuller gas diffusivity ⇒ Lewis number (wet-bulb).

The caloric engine: H, S, G from C_p

This is the mechanism behind "other properties that derive from here" — and it is worth seeing *how one arrives at* $S(T)$, since that is the step that trips everyone up. Start from the laws, at constant pressure. Heat capacity *is* the

temperature slope of enthalpy, $C_p = (\partial H/\partial T)_P$, so

$$dH = C_p\,dT.$$

For entropy, the second law gives $dS = \delta q_{\text{rev}}/T$; at constant pressure the reversible heat is $\delta q_{\text{rev}} = C_p\,dT$, hence

$$dS = \frac{C_p}{T}\,dT.$$

That single step — dividing by T — is the whole origin of the $\int C_p/T$. Integrate each from a reference temperature and add the 298 K anchor; Choupo stores $C_p^\circ(T)$ and the two anchors (*not* tables of H, S) and evaluates:

$$H^\circ(T) = \underbrace{\Delta H_f^\circ}_{\text{position}} + \int_{298.15}^{T} C_p^\circ(T')\,dT' \qquad (\texttt{Component::h\_pure\_ig},\ \texttt{Component.cpp}) \tag{1}$$

$$S^\circ(T) = \underbrace{S_{298}^\circ}_{\text{absolute}} + \int_{298.15}^{T} \frac{C_p^\circ(T')}{T'}\,dT' \qquad (\texttt{Component::s\_pure\_ig}) \tag{2}$$

$$G^\circ(T) = H^\circ(T) - T\,S^\circ(T) \tag{3}$$

For a mixture the engine adds the entropy of mixing $-R\sum y_i \ln y_i$ and the pressure term $-R\ln(P/P_{\text{ref}})$ (`ThermoPackage::S_ig`). That is the *entire* ideal-gas caloric model — no hidden tables.

These integrals are not done numerically. For a polynomial C_p the primitive is elementary — it was solved *once, on paper* — so each call evaluates a closed form,

$$\int_{T_{ref}}^{T} \frac{C_p^\circ}{T'}\,dT' = a_0 \ln\frac{T}{T_{ref}} + \sum_{k\geq 1}\frac{a_k}{k}\left(T^k - T_{ref}^k\right),$$

exact to machine precision in a handful of operations (`PolynomialCp::S`). The standard correlations all share this property (NASA, Aly–Lee, Shomate have closed primitives too); Choupo never runs a quadrature loop for C_p.

Worked example 1.1. Acetone H and S at 400 K — by hand, then by ChoupoWith $C_p^\circ(T) = 6.301 + 0.2606\,T - 1.253\times10^{-4}T^2 + 2.038\times10^{-8}T^3$ (units J mol^{-1}K^{-1}), integrating from 298.15 to 400 K:

$$\int C_p^\circ\,dT' = 8431\ \text{J/mol}, \qquad \int \frac{C_p^\circ}{T'}\,dT' = 24.19\ \text{J/(mol}\cdot\text{K)}.$$

So $H^\circ(400) = -217100 + 8431 = \mathbf{-208669}$ J/mol and $S^\circ(400) = 295.30 + 24.19 = \mathbf{319.49}$ J/(mol · K). Running `choupoProps` on a one-point scan returns exactly $H_{\text{ig}} = -208669$ J/mol, $S_{\text{ig}} = 319.49$ J/(mol · K). The number on the page is the number in the code — and at 298.15 K both integrals vanish, so $H = \Delta H_f^\circ$ and $S = S_{298}^\circ$, the anchors laid bare.

Key idea. C_p gives the *shape*; the formation values give the *position*. That is why the `.dat` lists $C_p^\circ(T)$ *and* ΔH_f°, S° separately — not redundancy. And note the asymmetry: entropy has an absolute zero for free (third law: $S\to0$ for a perfect crystal at $T=0$), so S_{298}° is a genuine absolute; enthalpy has *no* natural zero, so ΔH_f° must be *measured* and supplied.

What the critical properties and acentric factor actually do

T_c, P_c, ω do nothing in the caloric equations above — they enter only when you ask for *real-fluid* behaviour through a cubic equation of state. There the attraction term is $a(T) = a_c\,\alpha(T_r)$, with a_c fixed by T_c, P_c and the temperature shape fixed by ω (Soave/SRK):

$$\alpha(T) = \left[1 + m(1-\sqrt{T_r})\right]^2, \quad m = 0.48508 + 1.55171\,\omega - 0.15613\,\omega^2, \quad T_r = \tfrac{T}{T_c}$$

(`SRK.cpp`; Peng–Robinson uses the analogous $\kappa(\omega)$). This $a(T)$ is what produces the departure functions $H - H^\circ$, $S - S^\circ$ and the compressibility Z — the real-gas correction layered on top of the ideal-gas caloric result. The same T_c, T_b also drive Watson's latent heat, $\Delta H_{\text{vap}}(T) = \Delta H_{\text{vap}}(T_b)\,[(T_c - T)/(T_c - T_b)]^{0.38}$ (`Component.cpp`).

> *Intuition.* So why does `acetone.dat` carry ω at all if you are running the default `idealGas`? Because the data is *model-agnostic* and the model decides which fields wake up. Under `idealGas` there is no departure, and ω, T_c, P_c are simply never read. Select `SRK` for a high-pressure compressor and the very same numbers come alive. The `.dat` is the laboratory record; the model is the question you ask of it.

1.6 From G to equilibrium: what ΔG° is (and is not)

The single most common confusion in this whole subject is ΔG° versus ΔG. It is worth settling once, because G is what the reactors minimise and it is built from exactly the H and S above.

First, the $^\circ$. It means *reference state* — each species pure, ideal gas, at 1 bar — but *at the reaction temperature* T. It does *not* mean "at 298 K." So $\Delta G^\circ = \Delta G^\circ(T)$ varies with temperature; $\Delta G^\circ(298)$ is just one point on a curve.

Then, the two are not the same thing.

- ΔG° is the Gibbs change if reactants → products with *everything in its standard state*. It is a *fixed number* — a property of the reaction at T.
- ΔG is the *actual* Gibbs change at the real composition (the real partial pressures). It is the *driving force*.

The reaction quotient Q bridges them, and equilibrium is where the driving force — not the reference number — vanishes:

$$\Delta G = \Delta G^\circ + RT\ln Q, \qquad \Delta G = 0 \text{ at equilibrium } \Rightarrow \boxed{\Delta G^\circ = -RT\ln K}$$

> **Key idea.** At equilibrium it is ΔG that is zero, *never* ΔG°. ΔG° stays fixed and merely sets *where* the equilibrium sits (the value of K); the system gets there by shifting Q until $\Delta G = 0$. And $\Delta G^\circ > 0$ does not mean "no reaction" — it means $K < 1$, i.e. partial conversion. Only the sign and size of K change.

Where the number comes from — straight back to the caloric engine:

$$\Delta G^\circ(T) = \Delta H^\circ(T) - T\,\Delta S^\circ(T) = \sum_i \nu_i\, g_i^\circ(T), \qquad g_i^\circ(T) = h_i^{\mathrm{ig}}(T) - T\, s_i^{\mathrm{ig}}(T).$$

Note the subtlety that confuses everyone: h^{ig} carries the *formation* enthalpy (ΔH_f°, an elements datum) while s^{ig} is the *absolute* (third-law) entropy. Mixing the two references looks wrong — but it is exactly right, because in a *balanced* reaction the element contributions cancel through $\sum_i \nu_i$. That is precisely why the `.dat` carries ΔH_f° *and* S° separately: each plays a different role, and they combine correctly only across a balanced stoichiometry (`Component::g_pure_ig`; consumed in `PFR.cpp`/`CSTR.cpp` as $K = \exp(-\Delta G^\circ/RT)$ and in the Gibbs reactor's direct G-minimisation).

How K moves with T (Gibbs–Helmholtz). $\Delta G^\circ(T)$ is not a frozen datum — it has its own temperature law, and that law is a workhorse relation worth knowing by name. Gibbs–Helmholtz states $[\partial(G/T)/\partial T]_P = -H/T^2$; applied to $\Delta G^\circ = -RT\ln K$ it collapses to the van 't Hoff equation,

$$\boxed{\frac{d\ln K}{dT} = \frac{\Delta H^\circ}{RT^2}}$$

which is just Le Chatelier made quantitative: an exothermic reaction ($\Delta H^\circ < 0$) has K *falling* as T rises. In Choupo it doubles as a glass-box self-check — the $K(T)$ the reactors build from $g^\circ(T)$ at each temperature must show exactly this slope, or the enthalpy and the equilibrium constant are disagreeing with each other.

> *Intuition.* These relations are not ornaments. Gibbs–Helmholtz is *how the equilibrium moves with temperature* (above); its sibling, the **Gibbs–Duhem** equation $\sum_i x_i\, \mathrm{d}\ln\gamma_i = 0$ (constant T, P), says the activity coefficients of a mixture are *not independent*. That single constraint is exactly why measured VLE data can be *tested* for thermodynamic consistency — and why an activity model is built from one G^{E} function rather than fitting each γ_i freely. It is the engine of §8.

1.7 A worked span: water / ethanol / benzene, -200 to 1000 °C

Put the three together and sweep the temperature. This one mixture crosses *every* regime of property prediction, and is the honest way to see both how properties are obtained *and* where the models give out — which is exactly the colour band of Figure 2, now read regime by regime. (Critical temperatures: ethanol 241 °C, benzene 289 °C, water 374 °C; melting points: ethanol −114, water 0, benzene 5.5 °C.)

Regime	How a property is obtained	Where the model gives out
-200 to $\sim 0\,^\circ$C solids (SLE)	melting points T_m, ΔH_{fus}, solid C_p; freezing-point depression; sublimation P^{sat}	**low-T edge.** Below every Antoine `Trange`: P^{sat} would extrapolate blindly. Solid-phase data is sparse — a cryogenic *data desert*.
~ 0 to $\sim 65\,^\circ$C two liquids (LLE)	activity model (`NRTL`/`UNIFAC`) for γ; water/benzene nearly immiscible, ethanol the cosolvent	VLE-fitted parameters do *not* serve LLE; `Wilson` cannot represent a liquid split at all; without T-dependent parameters the miscibility gap appears/vanishes wrongly.
~ 65 to $\sim 100\,^\circ$C boiling (VLLE)	Antoine P^{sat} + γ–ϕ + three-phase flash (Gibbs minimisation); ethanol/water + ternary azeotrope ($\sim 64.9\,^\circ$C)	azeotrope sensitivity — a small γ error moves the column design; bubble-point (Wang–Henke) distillation is unstable *through* an azeotrope (use `method simultaneous`).
$\sim$ 100 to $\sim$ 290 °C vapour → near-critical	cubic EoS (`SRK`/`PR`) with k_{ij}; departure functions $H - H^\circ$, $S - S^\circ$, Z	cubics degrade as the critical region is approached; k_{ij} often needs to be made temperature-dependent over so wide a span.
~ 290 to $374\,^\circ$C mixed super/sub-critical	the method itself must switch γ–$\phi \to \phi$–ϕ (one EoS for every phase)	above a species' T_c the activity-coefficient picture loses meaning; no single γ model spans the switch; near-critical accuracy is poor.
374 to $1000\,^\circ$C supercritical + reaction	ϕ–ϕ EoS; wide-range $C_p^\circ(T)$ for the caloric terms	**high-T edge.** P^{sat} ceases to exist above T_c (Choupo *refuses* it, §1.2); and the species thermally *decompose* — the fixed-component assumption fails: do they still exist?

The two edges — the ones you asked to watch — fail for opposite reasons but teach the same lesson. At the **cold edge** there is simply no data: every correlation is fitted to a window that starts well above 73 K, the species are frozen, and a confident number there would be invention, not prediction. At the **hot edge** the theory keeps producing numbers — Antoine will gladly report a vapour pressure at 800 °C — but the *physics* has moved on: past each T_c no saturation exists, and past a few hundred degrees the molecules are no longer the molecules in the `.dat`. In both cases the correct output is a *flag*, not a figure.

Common pitfall. No single model spans this axis. A run that quietly uses one γ model and one cubic from -200 to 1000 °C is not "general" — it is extrapolating through three regime changes in silence. The wide span is not one problem; it is five problems wearing one temperature axis.

Intuition. What Choupo ships *today* covers the middle of this span honestly: cubic EoS, `ideal`/`NRTL`/`Wilson`/`UNIFAC` activity, Antoine P^{sat}, Watson ΔH_{vap}, three-phase VLLE by Gibbs minimisation, and azeotropic columns via the simultaneous MESH solver. It does *not* (yet) carry association models (PC-SAFT/CPA) for the hydrogen-bonding non-ideality, nor electrolyte or wide-range near-critical packages. So for the H-bonding mid-span and the two edges, Choupo's job is not to fake accuracy it does not have — it is to predict where it can and to *say so* where it cannot. That honesty is the product.

2 The Choupo property architecture — the complete map

Everything in this guide hangs off one architecture. This chapter is the complete map, stated once; every later chapter is a room in this house.

2.1 The flow

`constant/thermoPhysPropDict` *declares* → the builder *loads, verifies and announces* (it never estimates) → the engine *computes* → the run log *confesses*.

Estimation is a curation-time act (the props bench writes a reviewable `.dat`); assembly-time refusal is loud (a missing declared parameter file, a k_{ij} regressed for another cubic, a vapour method the liquid path cannot serve). The log then shows the whole assembly: the VLE world, the reference rung of every group, and every pair parameter with its source file.

2.2 The grammar — the axes of a property package

Axis	Slot	What it
components	`components ( ... )`	identity; a ions
phases	`phases`, `phases/solid` data	which pha
model per PHASE	`equilibrium { formulation ...; liquid {...} vapour {...} }`	ONE Gib `formulati`
convention per GROUP	reference rungs (`referenceBasis`)	renormalis Raoult, so infinite dil
correlation per COMPONENT	the component `.dat` (+ overrides)	pure-speci pure corre surface
parameter per PAIR	`parameters { ...Pairs {...} }`	declared F to its sour
mixing per PROPERTY × PHASE	`mixing` slots + `effective{}`	how pure Nissan, va closures (a viscosity) l are *media* ing
chemistry	`chemistry { salts (...); }`	real equili
package per UNIT	unit `thermo {}`	the spatial whole pac single axes in one pha dited (ΔH
phase equilibrium	`phaseEquilibrium {}`	for each imposed DATUM `datum aqu` `<pair-fil` surfaces, implicit

Provenance is an *attribute of every value*, not an axis: each resolved number carries a quality word (`measured > fitted > correspondingStates > estimated > asserted`), and an `asserted` value — a pinned constant, a user polynomial, a user table — is refused without its written `why`.

2.3 The five formulations

A case declares one `equilibrium.formulation`; the builder assembles that one natively and gives every other shape a named refusal. Each is validated against measured data in this guide:

Formulation	Liquid method	K-value	Natura
`gammaPhi`	`activity.NRTL` (Wilson, UNIQUAC, UNIFAC, cosmoSAC)	$\gamma_i P_i^{sat}/P$	organic
`gammaGamma`	one activity model *per named liquid phase*	$\gamma_i^{\alpha}/\gamma_i^{\beta}$	LLE / V
`diluteSolution`	`solution.henryDilute`	$\gamma_i^* H_i/P$	dissolved
`phiPhi`	`eos.SRK` / `eos.PengRobinson` / `PCSAFT`	φ_i^L/φ_i^V	high pre
`electrolyteGammaPhi`	`Pitzer` / `eNRTL` (molality)	ion activities, osmotic	brines, s

`gammaGamma` is the one that surprises people: two coexisting liquids are *two phases*, each entitled to its own Gibbs surface, and the split is found by minimising G directly rather than by iterating K — a symmetric γ-model would otherwise collapse onto the trivial $K = 1$ saddle. That is the legality rule of §3 doing visible work.

2.4 The five component routes

1. **Standard, by name** — `data/standards/components/` (exact-name $O(1)$, every value cited, the engine refuses to write there).
2. **Local, by name** — `data/local/components/` is the *private* working tier (gitignored; shipped empty). Records are runtime-usable, but every use is announced as `[local] UNVERIFIED`; a local record never shadows a standard one. Integrity and provenance gates make a local record inspectable, not automatically authoritative.
3. **New, defined in the case** — `myCase/constant/components/`: a new molecule (estimated and labelled), a new ion, a pseudo-component lump. The case runs self-contained from day one; promotion into the standards tree is a *requested, reviewed curation act* after validation.
4. **Standard + overlay** — a partial case file merges field-by-field over the standard entry (sample-specific measured data, a new model block); the run announces the merge.
5. **Solids** — solid reference rungs, solids on their own stream channel, and the iron rule: a dissolving salt's solid formation enthalpy is ION-DERIVED ($H_f^{solid} = \sum \nu_i h_{f,aq,i} - \Delta H_{soln}$), never a stored duplicate.

Thus component resolution is **case-local overlay > standard > local (private, announced)**. This is an evidence ladder, not merely a search path: moving a file up the ladder requires review and validation, not a rename.

2.5 The ChemSep import route

The ChemSep 8.3 importer (`bin/curate/chemsep_to_choupo.py`) is deliberately reproducible and reviewable. It runs CAS-based de-duplication against the standard catalogue, then a structural/provenance audit and an internal $P^{\mathrm{sat}}(T_b)$ closure check on every imported component record — a closure check on the imported constants and predictive correlation, not an independent comparison with laboratory data.

The same import stages NRTL, UNIQUAC and Wilson binary-pair records, each passing identity, direction, licence, source-digest and exact unit-conversion checks. Their source does not provide a captured temperature-validity range, so they remain screening data and the runtime prints `[local] UNVERIFIED` when one is selected. Pair resolution is **case-local/inline > standard > local > announced ideal default**; local data can therefore extend a case without silently changing an approved standard.

The source boundary is intentional. The unmodified third-party database and its Artistic-2.0 licence are kept under `thirdParty/chemsep/` (gitignored); the normalised, attributed records are written to the *private* `data/local/` tier — they are never redistributed with the public repository. Literature values retain literature provenance; Ambrose–Walton and other closures retain predictive provenance. Missing C_p, formation, electrolyte, adsorption, kinetic or transport data are not inferred merely to make a record look complete.

The case additionally owns its *equipment physics*: reaction kinetics in `constant/reactions` (the heat of reaction always computed from the formation datum and announced), per-machine measured curves (a dryer's falling-rate data) in the unit's own folder, all inherited through the LOUD fractal cascade (plant → sector → unit).

2.6 The reference case

The whole architecture is exercised by one designed flagship, `tutorials/plant/lithiumBrinePlant` (under construction): salar brine through Pitzer ponds and crystallisers, NF membrane Mg/Li separation, solvent extraction with a student-defined extractant (the γ–φ world and the audited Pitzer↔NRTL seam), electrodialysis, 60-bar CO_2 carbonation (the φ–φ override and the Henry world), a radical-kinetics burner satellite for thermal NO_x, batch recrystallisation recipes, a control satellite, and the full outer ring (fitting, sweeps, optimisation, sizing, costing, economics). Later chapters refer to its sectors by name.

2.7 The principle

No silent crutch, anywhere: every value is cited, estimated-and-labelled, overlaid-and-announced, or asserted-with-a-why — and the run log is where the whole assembly confesses. There is no ultimate model; hunting with the cat is honest engineering, hiding the cat is fraud.

3 The thermodynamic chain: how every equation here is derived

Every K-value, activity coefficient and Henry constant in this guide falls out of ONE chain of ideas. This section walks it once, so that no equation below ever has to be taken on faith. (Full derivations: the Theory Guide, chapters on flash, activity models, electrolytes and Henry's law.)

3.1 1. Equilibrium is a Gibbs minimum

The second law says an isolated system maximises entropy; recast at the conditions a process engineer actually controls — fixed T and P — it says a system minimises its **Gibbs energy** G. For a multicomponent, multiphase system,

$$dG \;=\; -S\,dT + V\,dP + \sum_i \mu_i\,dn_i, \qquad \mu_i \equiv \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}},$$

and at the minimum, moving a mole of i between phases must change nothing:

$$\boxed{\mu_i^{\alpha} = \mu_i^{\beta}} \quad \text{for every component, across every pair of phases.}$$

That single boxed statement is ALL the physics; everything else is bookkeeping to make it computable.

3.2 2. Fugacity: the workable currency

μ_i is awkward: its absolute value is unknowable and it dives to $-\infty$ as $x_i \to 0$. Lewis (1901) defined the **fugacity** f_i through $\mu_i = \mu_i^{\circ} + RT\ln(f_i/f_i^{\circ})$ — an "effective pressure" that equals y_iP in an ideal gas. Because ln is monotonic, the equilibrium condition becomes

$$\boxed{f_i^{\,\alpha} = f_i^{\,\beta}}$$

— equal fugacities, phase by phase, component by component. Every VLE "world" in Choupo is just a different, self-consistent way of WRITING f_i for a phase.

3.3 3. Two routes to a mixture fugacity

Route A — an equation of state (works for any phase the EoS can represent). Define the fugacity coefficient $\varphi_i \equiv f_i/(y_iP)$; thermodynamics gives it exactly from the volumetric behaviour,

$$RT\ln\varphi_i \;=\; \int_V^{\infty}\left[\left(\frac{\partial P}{\partial n_i}\right)_{T,V,n_j} - \frac{RT}{V}\right]dV \;-\; RT\ln Z,$$

so ONE model of $P(T,V,\mathbf{n})$ — a cubic like SRK with its $a(T)$, b and the pair correction k_{ij} — yields φ_i for BOTH the vapour root and the liquid root of the same cubic.

Route B — an activity model (liquids only). Split the liquid fugacity into a reference times a correction, $f_i^L = \gamma_i\,x_i\,f_i^{\circ}$, where the **activity coefficient** γ_i comes from a model of the excess Gibbs energy (G^E: NRTL, UNIQUAC, Wilson, Pitzer, eNRTL) via $RT\ln\gamma_i = \partial G^E/\partial n_i$. The price of Route B is choosing the **reference** f_i° — and that choice is a CONVENTION, not physics:

Convention	Reference f_i°	Normalised so
symmetric (Raoult)	pure liquid: P_i^{sat} ($\times\,\varphi_i^{sat}$·Poynting at high P)	$\gamma_i \to 1$ as $x_i \to 1$
unsymmetric (Henry)	infinite dilution: $H_{i,\mathrm{solv}} = \lim_{x\to 0} f_i/x_i$	$\gamma_i^{*} \to 1$ as $x_i \to 0$
electrolyte (aqueous)	ion at infinite dilution, $\mathrm{H}^+(\mathrm{aq}) = 0$ datum	$\gamma_{\pm} \to 1$ as $m \to 0$

A supercritical solute HAS no pure liquid, so its only definable anchor is the Henry limit — which is why dissolved H_2/N_2 take the unsymmetric convention, never an Antoine curve extrapolated above T_c.

3.4 4. One Gibbs surface per phase — the legality rule

Gibbs–Duhem, $\sum_i x_i\,d\ln\gamma_i = 0$ at constant T, P, ties every component's γ in a phase to every other's: they must all derive from ONE G^E (or one EoS) surface. Hence the rule that governs Choupo's `thermoPhysPropDict` grammar: *models per phase; conventions per component-group within the phase; correlations per component; parameters per pair.* Mixing two γ-models (or two cubics) in one phase violates Gibbs–Duhem and is refused. Mixing CONVENTIONS in one phase is legal — solvent symmetric, solutes unsymmetric — because a convention only renormalises the same surface.

3.5 5. The four worlds' K-values, assembled

Setting $f_i^L = f_i^V$ with each route/convention gives every K-value this guide uses:

world	f_i^L	f_i^V	$K_i = y_i/x_i$
γ-φ	$\gamma_i x_i P_i^{\text{sat}}$	$y_i\varphi_i P$	$\dfrac{\gamma_i\, P_i^{\text{sat}}}{\varphi_i\, P}$
Henry (KK/KI)	$\gamma_i^* x_i H_i(T)\, e^{\bar{v}_i^{\infty}(P-P_s)/RT}$	$y_i\varphi_i P$	$\dfrac{\gamma_i^* H_i\, \mathcal{P}}{\varphi_i P}$
φ-φ	$x_i\varphi_i^L P$	$y_i\varphi_i^V P$	$\dfrac{\varphi_i^L}{\varphi_i^V}$
electrolyte	ion activities from Pitzer/eNRTL on the aqueous rung; full speciation (mass action + charge balance) where the case demands it		

Each row is one `equilibrium.formulation` (`gammaPhi`, `diluteSolution`, `phiPhi`, `electrolyteGammaPhi`); the run log announces which world is live. All four are validated in the tutorials against measured data — van Ness-consistent VLE sets (γ-φ), Wiebe/Larson solubilities (Henry, $\sim$2%), the Stryjek N_2–CH_4 isotherm (φ-φ, $\sim$2%), and the Hamer–Wu / calorimetric electrolyte sets.

> *Intuition.* The whole edifice is: G minimum → equal μ → equal f → pick HOW to write f per phase (an EoS or a G^E model + a reference convention) → a K-value. When an equation in this guide looks arbitrary, walk it back up this chain — it never is.

4 Why a separate guide

The Theory Guide answers *how a model works* — how a flash splits, how an EoS predicts Z, how Levenberg–Marquardt minimises a residual. This guide answers a question that comes *before* that: *where do the numbers that feed the models come from?* A simulation is only as good as the T_c, the NRTL pair, the heat of formation it stands on, and those are not handed down — they are estimated, fitted, and tested.

Choupo treats that as a first-class phase with its own binary (`choupoProps`) and its own lifecycle:

> **CREATE** a component (from groups) → **CURATE** the mixture models (fit pairs, test data, choose by evidence) → **WIRE** the flowsheet → **SIMULATE**.

A `propsDict` of `operations` drives the first two stages. A case may have *no* `flowsheetDict` at all (a pure properties case — the compound bench), or carry one and use props to curate the thermo it will simulate. Same files, same engine; only the centre of the GUI differs.

> *Intuition.* The usual reflex is to click a property method and trust a badge. Choupo's stance (the "see, then decide" principle) is the opposite: you make the estimate or the fit *visible* — the group build-up, the model curve against the lab points, the consistency residual — and decide from the evidence in front of you. Estimation has error; the point is to see it.

5 Where curated data lives, and how a case selects it

You estimated a T_c, fitted an NRTL pair, tested a heat of solution. Two questions follow: *where does that number land*, and *how does a case use it*? Choupo's answer is *declarative property packages with OpenFOAM-like explicit files*.

A curated value lands in a glass-box `.dat` in a tree organised by *what kind of thing* it is — the intrinsic property of one substance, the parameter of a pair, the equilibrium of a reaction:

```
data/standards/
  components/<name>.dat            intrinsic, one substance (Tc, Antoine, Cp)
                                   + dissociatesTo { Na 1; Cl 1; }  [salts]
  species/<name>.dat               one modelSpecies file per species
                                   (Na, Cl, CO2aq, O2): formula + charge +
                                   medium-tagged thermo; never fed to a flowsheet
  phases/solid/<phase>.dat         crystal phase (rho_p, k_v): halite, sylvite
  chemistry/.../<name>.dat         an equilibrium: dissolution, speciation
  parameters/.../<i>-<j>.dat       a pair: NRTL, Pitzer, eNRTL
  methods/<name>.dat           a model declaration (+ its reference basis)
  (the system is declared INLINE in the case's constant/thermoPhysPropDict
   -- THE CENTRE; there is no shared package catalogue)
```

The arity rule decides *which* file: a number that "stands alone with the molecule" (the T_c of water) lives with the component; a number that needs two names (an NRTL pair) cannot, and goes to `parameters/`. The Developer Guide (*the thermophysical property architecture*) gives the full layout and the eight data kinds.

Declaring the system. A case does not list loose models and parameters. It *declares* its thermophysical system — one record that names the components, the formulation, and the parameter files the run needs (active chemistry lives beside it in `constant/chemistryDict`):

```
recordType    thermophysicalPropertySystem;
schemaVersion 2;

components ( water NaCl );          // NaCl's dissociatesTo names the ions

equilibrium
{
    formulation electrolyteGammaPhi;

    aqueous
    {
        solvent            water;
        apparentComponents ( NaCl );  // stream carries the salt; the model
        activityModel { model Pitzer; }        //   activates its ions
        compositionBasis molality;
    }

    vapour { fugacityModel idealGas; }
}
% Pitzer pairs resolve by (cation,anion) name from parameters/; declared
% binaryParameters/binaryInteractions files load from their 'source' paths
% and refuse when missing.
```

The flow is one sentence: **`thermoPhysPropDict` declares, `ThermoPackageBuilder` assembles, `ThermoPackage` computes.** The declaration is the recipe; the builder loads what it names and mounts the runtime object the unit operations consume. An ordinary molecular case is the same record degenerate — no dissociation, no chemistry, a `gammaPhi` block with an NRTL or an EoS slot.

> *Intuition.* The boundary you must respect: *curation estimates, runtime does not.* You may estimate, fit, and derive at the props bench — and you *see* it (§6, §7). Once a case runs, the builder only *loads* existing case, standard or local records; it never invents a missing property. For activity models, an unresolved binary

pair may use the documented ideal default, but that fallback is announced and appears in the provenance. A required record with no defined fallback stops the run with a clear message. So: estimate, look, promote — and inspect every local or defaulted rung used by the package.

6 Creating a component: Joback group contribution

When a component is missing from the catalogue and from the literature, you can *estimate* its pure-component constants from its molecular structure — the groups it is built from. Choupo implements the **Joback** method (Joback & Reid, 1987; tabulated in Poling, Prausnitz & O'Connell, *The Properties of Gases and Liquids*, 5th ed.), the `estimateComponent` property operation.

6.1 The method

Each first-order group contributes a tabulated increment. Summing the increments and substituting into Joback's correlations gives the constants (with n_A the total number of atoms):

$$T_b = 198.2 + \textstyle\sum_i n_i\,\Delta T_{b,i} \tag{4}$$

$$T_c = T_b \,/\, \left(0.584 + 0.965\,\Sigma\Delta T_c - (\Sigma\Delta T_c)^2\right) \tag{5}$$

$$P_c = \left(0.113 + 0.0032\,n_A - \Sigma\Delta P_c\right)^{-2} \quad \text{[bar]} \tag{6}$$

$$V_c = 17.5 + \textstyle\sum_i n_i\,\Delta V_{c,i} \quad [\text{cm}^3/\text{mol}] \tag{7}$$

$$\Delta H_f^\circ(298) = 68.29 + \textstyle\sum_i n_i\,\Delta H_{f,i} \quad \text{[kJ/mol, ideal gas]} \tag{8}$$

$$\Delta H_{\text{vap}}(T_b) = 15.30 + \textstyle\sum_i n_i\,\Delta H_{v,i} \tag{9}$$

$$C_p^{\text{ig}}(T) = (\Sigma a - 37.93) + (\Sigma b + 0.210)T + (\Sigma c - 3.91\times10^{-4})T^2 + (\Sigma d + 2.06\times10^{-7})T^3. \tag{10}$$

The acentric factor ω is *not* a group sum; Choupo derives it from (T_b, T_c, P_c) by the Lee–Kesler correlation, with $T_{br} = T_b/T_c$ and P_c in atm:

$$\omega = \frac{-\ln P_c - 5.97214 + 6.09648/T_{br} + 1.28862\ln T_{br} - 0.169347\,T_{br}^6}{15.2518 - 15.6875/T_{br} - 13.4721\ln T_{br} + 0.43577\,T_{br}^6}. \tag{11}$$

A few first-order group increments (Joback & Reid; the full table is in the code, `src/propertyOps/EstimateComponent.cpp`):

group	ΔT_b	ΔT_c	ΔP_c	ΔH_f (kJ/mol)
$-\text{CH}_3$	23.58	0.0141	−0.0012	−76.45
$>\text{CH}_2$	22.88	0.0189	0.0000	−20.64
−OH (alcohol)	92.88	0.0741	0.0112	−208.04
>C=O (ketone)	76.75	0.0380	0.0031	−133.22
aromatic =CH−	26.73	0.0082	0.0011	2.09

6.2 The structure is identity: `jobackGroups` and `bin/estimate`

The molecular structure belongs *with the substance*, like the formula: a component's `.dat` declares its groups once,

```
jobackGroups { CH3 2; ketone 1; }     // acetone = 2 x CH3 + 1 x >C=O
```

(the sibling of the UNIFAC `groups{}` block — a different table, the same idea), and estimation needs no case at all:

```
bin/estimate acetone            # case-local first, then the standard catalogue
bin/estimate path/to/new.dat    # or an explicit file
```

The command runs the Joback engine on the component's own groups, prints the full glass-box build-up (each group's increment, the closure chain, deviations against any `reference` block), and drops *one* stable proposal — `<name>.estimated.dat` — next to the source, for you to review and promote by hand. It never writes into the frozen `data/standards/` tree, and the solver still refuses to estimate at run time: estimation is a *curation* act with a paper trail, not a silent fill-in.

The `estimateComponent` operation remains the pedagogical form — declare the groups inline in a `propsDict` when you want the estimate itself to be the lesson, as in the worked example below.

6.3 Worked example: acetone

Acetone, $CH_3-CO-CH_3$, decomposes into $2\times$ `CH3` and $1\times$ `ketone`. The operation prints the build-up and validates against a `reference` block:

Property	Joback	Reference (NIST)	Dev.
T_b	322.1 K	329.2 K	−2.2%
T_c	500.6 K	508.1 K	−1.5%
P_c	48.0 bar	47.0 bar	+2.2%
V_c	209.5 cm^3/mol	209 cm^3/mol	≈ 0
ω	0.295	0.307	−3.9%
ΔH_f°	−217.8 kJ/mol	−217.1 kJ/mol	+0.3%

Within a few percent on the criticals and almost exact on ΔH_f — and you *see* the error, because the reference sits beside the estimate.

Intuition. **Know the method's limits.** Joback is weak on T_b of strongly hydrogen-bonding species (ethanol's T_b comes out ∼14 K low), and it yields *no* Antoine / vapour-pressure parameters — those need a fit (§7) or a corresponding-states correlation, a separate step. An estimate is a starting point you then anchor to data, never a final answer.

Runnable case. `tutorials/props/estimate/estimate_acetone` — the example above, groups in, validated estimate out.

7 Curating mixture models: fitting binary pairs

A pure-component set is not enough for a mixture: the non-ideality lives in the *binary interactions* (NRTL, Wilson). Those parameters are regressed against experimental VLE data by Levenberg–Marquardt — the `fitParameters` operation.

7.1 The activity-coefficient models

For a binary, the **NRTL** model gives the liquid-phase activity coefficients from three adjustable numbers per pair — τ_{12}, τ_{21} (temperature-dependent) and the non-randomness α:

$$\tau_{ij} = a_{ij} + \frac{b_{ij}}{T}, \qquad G_{ij} = \exp(-\alpha\,\tau_{ij}), \tag{12}$$

$$\ln\gamma_1 = x_2^2\left[\tau_{21}\left(\frac{G_{21}}{x_1 + x_2 G_{21}}\right)^2 + \frac{\tau_{12}\,G_{12}}{(x_2 + x_1 G_{12})^2}\right], \tag{13}$$

and $\ln\gamma_2$ by swapping the indices. So a pair carries four fittable coefficients $(a_{12}, b_{12}, a_{21}, b_{21})$ plus a fixed α (often 0.3). **Wilson** is the two-parameter alternative,

$$\ln\gamma_1 = -\ln(x_1 + \Lambda_{12}x_2) + x_2\left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12}x_2} - \frac{\Lambda_{21}}{x_2 + \Lambda_{21}x_1}\right], \tag{14}$$

which cannot represent a liquid–liquid split (use NRTL when one is suspected).

7.2 Predicting without pairs: COSMO-SAC

NRTL and Wilson need a fitted pair — somebody measured that binary. **COSMO-SAC** is the predictive alternative: the activity coefficients come from each compound's *sigma profile* — the histogram of screening charge density over its molecular surface, from a quantum-chemical calculation done *once, elsewhere, at curation time* — and no binary parameter exists at all. Choupo implements the 2002 variant (Lin & Sandler), following the NIST benchmark implementation, as a plain activity model: same interface as NRTL, selected the same way.

A compound carries its profile in its own `.dat`, as one or more *named parameter sets* (a compound may carry profiles from several public collections, each labelled with its source):

```
cosmo
{
    VT2005            // set name = its source collection
    {
        variant       cosmoSAC2002;
        source        "Mullins et al., IECR 45 (2006) 4389";
        area          [-] 106.3;      // A^2
        volume        [-] 118.8;      // A^3
        sigmaProfile  ( ... );        // 51 points, -0.025..0.025 e/A^2
    }
}
```

The case selects the model and, when a compound carries more than one set, names which one:

```
activityModel { model cosmoSAC;  source VT2005; }
```

A compound with a single set needs no `source` (its lone set is the default); a multi-set compound without one is a loud error, as is a missing `cosmo` block or a set whose declared `variant` does not match the implemented constants — a profile is never silently combined with another variant's parameters. The standard catalogue ships VT2005 sets for 77 components. Validation: a pure component gives $\ln\gamma = 0$ exactly, and the implementation is cross-checked bit-for-bit against an independent code on the same equations and profiles. What Choupo deliberately does *not* do is generate profiles (that is quantum chemistry, a curation act outside the simulator) or bulk-import thousands of compounds. Reference case: `tutorials/props/molecular/cosmoSAC01_water_ethanol`.

7.3 The residual the fit minimises

The measurable is the bubble temperature. At a fixed pressure P and liquid composition x, the bubble point is the T that closes modified Raoult,

$$\sum_i x_i\,\gamma_i(x,T)\,P_i^{\mathrm{sat}}(T) = P \tag{15}$$

(a Newton-1D in T; P_i^{sat} from Antoine). Levenberg–Marquardt then drives the pair coefficients $\theta = (a_{12}, b_{12}, a_{21}, b_{21})$ to minimise the sum of squared bubble-temperature residuals against the data,

$$\min_{\theta} \sum_{k} \left[T_{\mathrm{bub}}(x_k; \theta) - T_k^{\mathrm{exp}} \right]^2. \tag{16}$$

> *Intuition.* **A low residual is not a good fit.** At a *single* pressure $\tau_{ij} = a_{ij} + b_{ij}/T$ is nearly linear in both a_{ij} and b_{ij}/T over the narrow T range of one isotherm/isobar, so a and b are *correlated* — many (a, b) pairs fit equally well and the optimum is a valley, not a point. `fitParameters` forms $J^{\mathsf{T}}J$ at the optimum and reports the parameter correlation + confidence intervals so you SEE this; the Theory Guide's *Levenberg–Marquardt* and *What is fittable* chapters derive the diagnostic. Fit across pressures, or fix b, to break it.

7.4 See, then decide: the model overlay

Before fitting, look. Declare one scan `operation` per candidate model (ideal, NRTL, Wilson), emitting the $(x, T_{\mathrm{bubble}}, y_{\mathrm{eq}})$ trio, and an `experimental {}` entry that overlays them — and, with a `dataset`, draws the measured points on top. The student sees which model the data actually picks: for ethanol/water at 1 atm the ideal curve misses the azeotrope entirely while NRTL lands on it.

> **Runnable case.** `tutorials/props/compare/compare_vle_etoh_water` — ideal vs. NRTL bubble/dew curves overlaid with measured VLE points (Carey & Lewis, 1932).

7.5 Fit, then promote

When a model earns its keep, `fitParameters` regresses its coefficients to the data and writes a proposal to `constant/parameters/<model>/<pair>.fit-<date>.dat`. Promotion is a deliberate, visible act: you inspect the parity plot and the identifiability statistics (confidence intervals, parameter correlation) the fit reports — never a "good fit" badge — and copy the proposal into the case's `constant/` when you trust it.

8 Testing the data itself: consistency

Before fitting a model *to* data, test whether the data is even thermodynamically consistent — a fit to bad data is a polished wrong answer. The `vleConsistency` operation does this in three steps.

8.1 Step 1 — activity coefficients straight from the data

It does not assume a model. From each measured (x_1, y_1, T) at the run pressure P, modified Raoult gives the activity coefficient of each component directly:

$$\gamma_i = \frac{y_i\, P}{x_i\, P_i^{\mathrm{sat}}(T)}, \qquad P_i^{\mathrm{sat}}(T) \text{ from Antoine.} \tag{17}$$

These are *experimental* γ_i — the test now asks whether they obey thermodynamics.

8.2 Step 2 — the Gibbs–Duhem point test

The Gibbs–Duhem equation is the constraint every real mixture must satisfy at constant T, P:

$$\sum_i x_i \,\mathrm{d}\ln\gamma_i = 0 \quad\Longrightarrow\quad x_1 \frac{\mathrm{d}\ln\gamma_1}{\mathrm{d}x_1} + x_2 \frac{\mathrm{d}\ln\gamma_2}{\mathrm{d}x_1} = 0 \;\; \text{(binary)}. \tag{18}$$

The operation evaluates the left-hand side at each interior point (central differences in x_1) and reports its max and mean magnitude — the *point residual*. A consistent set keeps it near zero. (Isobaric data adds a small $-H^{\mathrm{E}}/RT^2 \cdot \mathrm{d}T/\mathrm{d}x$ term, dropped here; stated as a caveat, not hidden.)

8.3 Step 3 — the Herington area test

Integrate (18) across the whole composition range: for an isothermal set Gibbs–Duhem forces the net area of $\ln(\gamma_1/\gamma_2)$ to vanish. Herington's isobaric form compares that net area to the absolute area and corrects for the boiling-range:

$$D = 100\,\frac{\left|\displaystyle\int_0^1 \ln\frac{\gamma_1}{\gamma_2}\,\mathrm{d}x_1\right|}{\displaystyle\int_0^1 \left|\ln\frac{\gamma_1}{\gamma_2}\right|\mathrm{d}x_1}, \qquad J = 150\,\frac{T_{\max} - T_{\min}}{T_{\min}}, \tag{19}$$

(both integrals by the trapezoid rule over the sorted points). The set **passes** when $|D - J| < 10$. The GUI's consistency card plots $\ln\gamma_1$, $\ln\gamma_2$ and $\ln(\gamma_1/\gamma_2)$ vs x_1 with the net area shaded — you SEE the $\pm$ areas cancel, not just the number.

Finally the infinite-dilution coefficients $\gamma_1^{\infty}, \gamma_2^{\infty}$ (the dilute-end limits) are reported — the most stringent point of any model later fitted to the set.

> **Runnable case.** `tutorials/props/compare/compare_vle_etoh_water` — the `vleConsistency` op on the Carey–Lewis ethanol/water data (Herington $|D - J| = 4.6 < 10$, PASS).

This guide grows with the props pillar. Planned: a second group-contribution method (Constantinou–Gani) to compare against Joback; group contribution for liquid viscosity; a structure (SMILES) front-end for the compound bench.