

ANÁLISE DA GASEIFICAÇÃO COM VAPOR PARA PRODUÇÃO DE HIDROGÊNIO VIA MODELO SUBSTITUTO DE APRENDIZADO DE MÁQUINA

Victor Hugo Xavier Bernardes^a, Thaylane Da Rocha Bezerra^a, Jhonatan Nolasco Nascimento^{a,b},
Marcelo Moreira da Costa^b, Sarah Arvelos Altino^{a,*}

^a Universidade Federal de Uberlândia – Uberlândia – MG

^b Universidade Federal de Viçosa – Viçosa – MG

* sarah.arvelos@ufu.br

RESUMO

A co-gaseificação de biomassa e plásticos com vapor de água é uma alternativa promissora para a produção de hidrogênio, embora envolva elevada complexidade devido à interação entre variáveis operacionais e composicionais. Neste trabalho, foi desenvolvido um modelo substituto baseado em aprendizado de máquina (XGBoost), treinado a partir de dados gerados via simulação no software AVEVA PRO/II, para predição da fração molar de hidrogênio e da produção de gás. O modelo apresentou elevado desempenho preditivo ($R^2 > 0,95$), capturando adequadamente relações não lineares do processo. A análise das variáveis indicou que a produção de hidrogênio é fortemente influenciada pela temperatura e pelo teor de hidrogênio da alimentação, enquanto a produção de gás depende principalmente da composição elementar. Os resultados demonstram o potencial da abordagem para análise e otimização de processos de gaseificação.

Palavras-chave: gaseificação, vapor d'água; hidrogênio; aprendizado de máquina; XGBoost

1. Motivação

A gaseificação de biomassa tem sido amplamente investigada como rota para produção de hidrogênio sustentável, sendo a co-gaseificação com plásticos uma alternativa promissora para aumentar o poder calorífico e a razão H/C da alimentação. Nesse contexto, a utilização de vapor de água como agente gaseificante favorece reações de reforma e deslocamento gás-água, resultando em maiores frações de hidrogênio no gás de (Song et al., 2022)..

Apesar dessas vantagens, o processo de co-gaseificação biomassa-plástico com vapor apresenta elevada complexidade, sendo governado simultaneamente por condições operacionais — como temperatura e razão vapor/biomassa — e pela composição físico-química da alimentação, expressa por variáveis como teores de carbono, hidrogênio e oxigênio. Essa combinação de fatores introduz forte não linearidade no sistema, dificultando sua análise por abordagens tradicionais baseadas apenas em modelos fenomenológicos (Janajreh et al., 2021).

Nesse cenário, técnicas de aprendizado de máquina têm sido aplicadas com sucesso na modelagem de processos de gaseificação, permitindo capturar relações complexas entre variáveis operacionais, composição da alimentação e respostas do sistema (Ascher et al., 2022). No entanto, a integração entre simulação fenomenológica e modelagem baseada em dados ainda é pouco explorada no contexto da co-gaseificação com vapor, especialmente considerando a variabilidade composicional dos materiais processados. Além disso, o uso do simulador AVEVA PRO/II para geração de dados nesse tipo de sistema permanece incipiente na literatura, apesar de sua capacidade de representar processos complexos por meio de balanços rigorosos de massa e energia (Bernardes et al., 2026).

Diante disso, este trabalho propõe o desenvolvimento de um modelo substituto baseado em aprendizado de máquina, treinado a partir de dados gerados via PRO/II, com o objetivo de analisar o comportamento da co-gaseificação biomassa-plástico com vapor, considerando simultaneamente condições operacionais e variáveis composicionais, e avaliar sua influência na produção de gás rico em hidrogênio.

2. Metodologia

2.1. Geração de dados via simulação (PRO/II)

Os dados utilizados neste trabalho (N=2993) foram gerados a partir de simulações do processo de co-gaseificação biomassa-plástico com vapor de água, utilizando o simulador comercial AVEVA PRO/II. Inicialmente, foram gerados 3000 dados e foi feita a remoção de um pequeno número de observações extremas nas variáveis de saída Gas Yield e H₂ (%), associadas a valores fisicamente inconsistentes, visando contribuir para

a melhoria da estabilidade da modelagem, sem comprometer a representatividade do conjunto de dados. O framework adotado segue a abordagem previamente proposta por (Bernardes et al., 2026) para gaseificação com ar, na qual o processo é representado por módulos sequenciais de secagem, pirólise, gaseificação e correção de equilíbrio restrito.

O modelo considera balanços de massa e energia sob condições estacionárias, com compostos não convencionais definidos a partir de análises elementares e imediatas da alimentação. A etapa de gaseificação é modelada por um reator de Gibbs, seguido por ajustes empíricos de equilíbrio para reações-chave, como water-gas shift e reforma do metano, visando reduzir desvios típicos de modelos de equilíbrio ideal.

A confiabilidade do modelo que foi apresentada por Bernardes et al. (2026). Para gaseificação com ar para foi previamente avaliada por comparação com dados experimentais da literatura para vapor de água considerando diferentes condições operacionais, apresentando erros médios (RMSE) na faixa de aproximadamente 2,8 a 5,0 pontos percentuais para as principais espécies do gás de síntese, indicando boa capacidade de reprodução das tendências globais do processo.

2.2. *Variáveis de entrada e saída, análise de redundância e pré-processamento*

As variáveis de entrada foram definidas de modo a representar simultaneamente as condições operacionais e a composição físico-química da alimentação. Foram consideradas como entradas:

- Composição elementar: C (%), H (%), O (%), N (%), S (%) e Cl (%);
- Análise imediata: cinzas (ASH), umidade, matéria volátil (VM) e carbono fixo (FC);
- Condições operacionais: vazão de vapor (STEAM em kg/h) e temperatura (°C).

Como variáveis de saída, foram selecionadas:

- fração molar de hidrogênio no gás de síntese (H_2 (%));
- produção de gás (Gas Yield em Nm^3/kg).

Inicialmente, foi realizada uma análise de redundância entre as variáveis de entrada por meio da correlação de Spearman, visando identificar dependências estruturais entre descritores composicionais. Pares de variáveis altamente correlacionadas foram avaliados individualmente, e a permanência no conjunto final considerou não apenas o grau de correlação, mas também a interpretabilidade físico-química e a relevância para a descrição do sistema.

2.3. *Modelagem por aprendizado de máquina e análise das variáveis*

A modelagem foi realizada utilizando o algoritmo eXtreme Gradient Boosting (XGBoost), amplamente empregado em problemas de regressão com dados tabulares devido à sua capacidade de capturar relações não lineares e interações entre variáveis (Chen & Guestrin, 2016). Os hiperparâmetros do modelo foram ajustados por meio de busca em grade (GridSearchCV) com validação cruzada de 5 partições, considerando diferentes combinações de profundidade das árvores, taxa de aprendizado, número de estimadores e frações de amostragem. O ajuste foi conduzido exclusivamente com os dados de treinamento, sendo o modelo final avaliado no conjunto de teste previamente definido.

O conjunto de dados foi dividido em subconjuntos de treinamento e teste na proporção de 80/20. Para mitigar efeitos de data leakage estrutural decorrentes da similaridade entre amostras, a partição foi realizada com base em uma estratégia de clusterização. Especificamente, foi aplicada a técnica de K-means ($k = 10$) sobre variáveis representativas da composição da alimentação, e os clusters obtidos foram utilizados como unidades de divisão, garantindo que amostras pertencentes ao mesmo cluster fossem alocadas integralmente em apenas um dos subconjuntos. O modelo foi ajustado com os dados de treinamento e avaliado em dados não vistos. O desempenho foi quantificado por meio do coeficiente de determinação (R^2) e do erro quadrático médio (RMSE), permitindo avaliar tanto a capacidade de ajuste quanto a generalização do modelo.

Após o treinamento do modelo, foi realizada uma análise da influência das variáveis de entrada sobre as saídas previstas, com o objetivo de identificar os principais fatores que governam o comportamento do sistema. Essa etapa foi conduzida de forma a relacionar os resultados obtidos com os mecanismos físico-químicos envolvidos na gaseificação com vapor, permitindo interpretar os efeitos observados à luz da engenharia de processos.

3. Resultados e Discussão

3.1. Visão geral do banco de dados e análise de correlação de Spearman

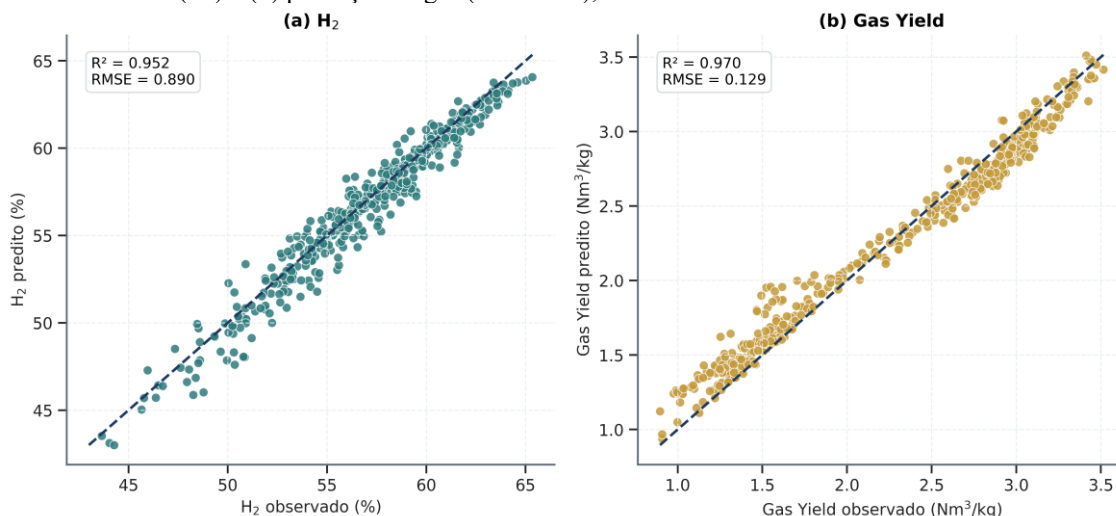
O banco de dados gerado utilizando o software PROII apresenta elevada diversidade composicional e operacional, abrangendo misturas biomassa-plástico ao longo de toda a faixa de composição (0 a 100% de plásticos), bem como ampla variação nas propriedades elementares e nas condições de processo (650–1000 °C e diferentes níveis de vapor). As variáveis de saída, em particular a fração molar de hidrogênio e a produção de gás, exibem distribuições contínuas e fisicamente coerentes. Os valores de H_2 obtidos (≈ 40 –68%) estão em consonância com resultados experimentais reportados para gaseificação com vapor, que tipicamente indicam concentrações de hidrogênio na faixa de 40–70% em volume (Luo & Chen, 2024). De forma semelhante, os valores de produção de gás ($\approx 0,77$ –4,43 Nm^3/kg) encontram-se dentro da ordem de grandeza esperada para sistemas de gaseificação de biomassa e misturas com plásticos (Thaha et al., 2026), reforçando a consistência físico-química do banco gerado.

A análise de correlação de Spearman revelou elevada redundância entre algumas variáveis de entrada. Em particular, observou-se correlação praticamente perfeita entre o teor de cinzas (ASH) e a matéria volátil (VM) ($\rho \approx 0,999$), indicando sobreposição de informação. Nesse caso, optou-se pela remoção da variável VM, mantendo-se ASH devido à sua maior interpretabilidade físico-química no contexto da gaseificação. Adicionalmente, foi identificada forte correlação entre os teores de carbono e oxigênio ($\rho \approx -0,93$), refletindo a relação inversa associada à composição elementar das misturas dessas variáveis na composição elementar da alimentação, especialmente em misturas biomassa-plástico. Diferentemente do caso anterior, ambas as variáveis foram mantidas, uma vez que representam aspectos distintos da estrutura química dos combustíveis e exercem papéis relevantes nos mecanismos de formação de gás. As demais variáveis foram preservadas, considerando sua contribuição para a descrição do sistema e a capacidade do algoritmo XGBoost em lidar com dependências entre descritores.

3.2. Desempenho do modelo

O desempenho dos modelos desenvolvidos é apresentado na Figura 1, por meio de gráficos de paridade entre valores observados e preditos para a fração molar de hidrogênio e a produção de gás. Para a predição de H_2 , o modelo XGBoost apresentou elevado desempenho no conjunto de teste ($R^2 = 0,952$; RMSE = 0,89 %), com boa aderência à linha de paridade e dispersão limitada dos pontos, indicando adequada capacidade de generalização. Para a variável Gas Yield, observou-se desempenho ainda superior ($R^2 = 0,970$; RMSE = 0,129 Nm^3/kg), com distribuição dos pontos fortemente concentrada ao longo da linha de paridade.

Figura 1. Gráficos de paridade entre valores observados e preditos para (a) fração molar de hidrogênio (H_2) e (b) produção de gás (Gas Yield), obtidos com o modelo XGBoost.



Observa-se que os coeficientes de determinação obtidos no conjunto de treinamento ($R^2 = 0,999$ para H_2 e $R^2 = 0,999$ para Gas Yield) permanecem superiores aos do conjunto de teste, indicando a presença de certo grau de sobreajuste (overfitting), ainda que controlado. No entanto, a consistência entre os resultados de validação

cruzada e teste sugere que o modelo mantém elevada capacidade preditiva quando aplicado a dados não vistos. Ressalta-se que o particionamento dos dados foi realizado com base em clusters, tornando a avaliação mais rigorosa ao evitar a presença de amostras altamente similares entre treino e teste.

3.3. Importância das variáveis de entrada

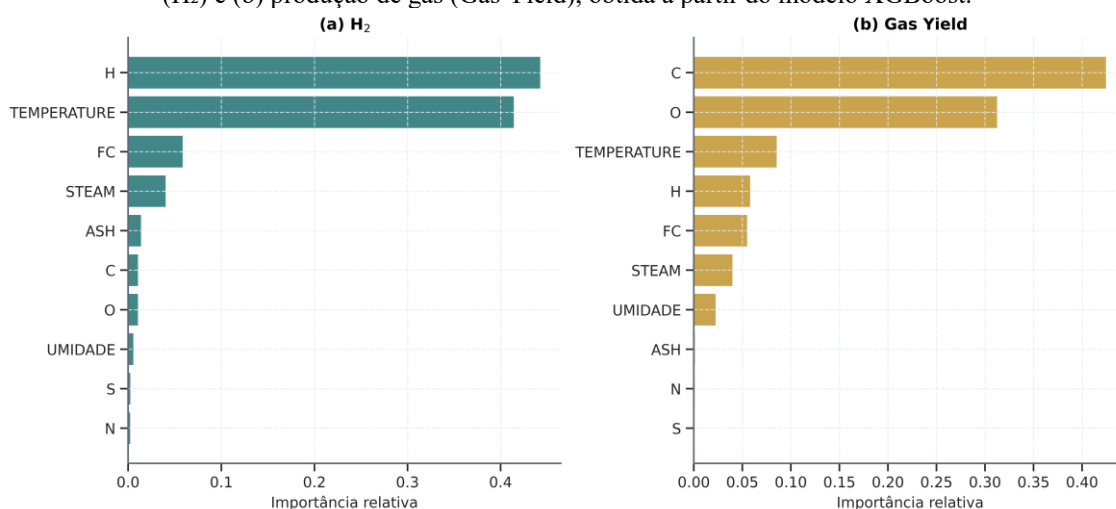
A Figura 2 mostra que a predição de H_2 foi dominada principalmente pelo teor de H da alimentação e pela temperatura, seguidos por contribuições bem menores de FC e STEAM. Esse resultado é fisicamente consistente: em gaseificação com vapor, o aumento da temperatura favorece reações endotérmicas, como reforma e gasificação do char, enquanto alimentações mais ricas em hidrogênio tendem a aumentar o potencial de formação de H_2 . Estudos experimentais e de revisão reportam que temperaturas mais elevadas e maior suprimento de vapor favorecem a produção de hidrogênio (Lu et al., 2014).

O papel de FC como variável relevante para H_2 também é plausível. Mesmo com importância inferior à de H e temperatura, o carbono fixo está associado à gasificação heterogênea do sólido residual, influenciando a formação de espécies intermediárias que participam das reações de reforma e deslocamento gás-água. Já a importância relativamente menor de STEAM no modelo, quando comparada à temperatura, sugere que, dentro da faixa operacional simulada, o efeito térmico é predominante, ainda que o vapor desempenhe papel importante na modulação das reações (Yalçinkaya et al., 2016).

Para Gas Yield, o padrão foi distinto. As variáveis mais importantes foram C e O, seguidas por temperatura, com contribuições secundárias de H, FC e STEAM. Esse comportamento é consistente com o fato de que a produção total de gás depende fortemente da composição global da alimentação. Em particular, o teor de carbono está diretamente associado ao potencial de geração de espécies gasosas, enquanto o teor de oxigênio influencia o grau de oxidação do combustível e sua tendência a formar produtos menos (De Las Obras-Loscertales et al., 2015).

De modo geral, os resultados indicam que as duas respostas analisadas não são controladas pelo mesmo conjunto de fatores. Enquanto H_2 é mais sensível ao potencial de hidrogenação da alimentação e à severidade térmica, Gas Yield é predominantemente governado pela composição elementar. Essa distinção é relevante do ponto de vista de engenharia, pois maximizar a fração de hidrogênio não implica necessariamente maximizar a produção total de gás. Além disso, a coerência entre as variáveis mais importantes identificadas pelo modelo e as tendências reportadas na literatura reforça a consistência físico-química do banco de dados e a adequação do modelo desenvolvido (Luo & Chen, 2024).

Figura 2. Importância relativa das variáveis de entrada na predição de (a) fração molar de hidrogênio (H_2) e (b) produção de gás (Gas Yield), obtida a partir do modelo XGBoost.



3.4. Análise de cenários operacionais

A Figura 3 apresenta a análise de cenários obtida a partir do modelo XGBoost, permitindo avaliar o efeito isolado das principais variáveis operacionais sobre as respostas do sistema. Para isso, as demais variáveis de

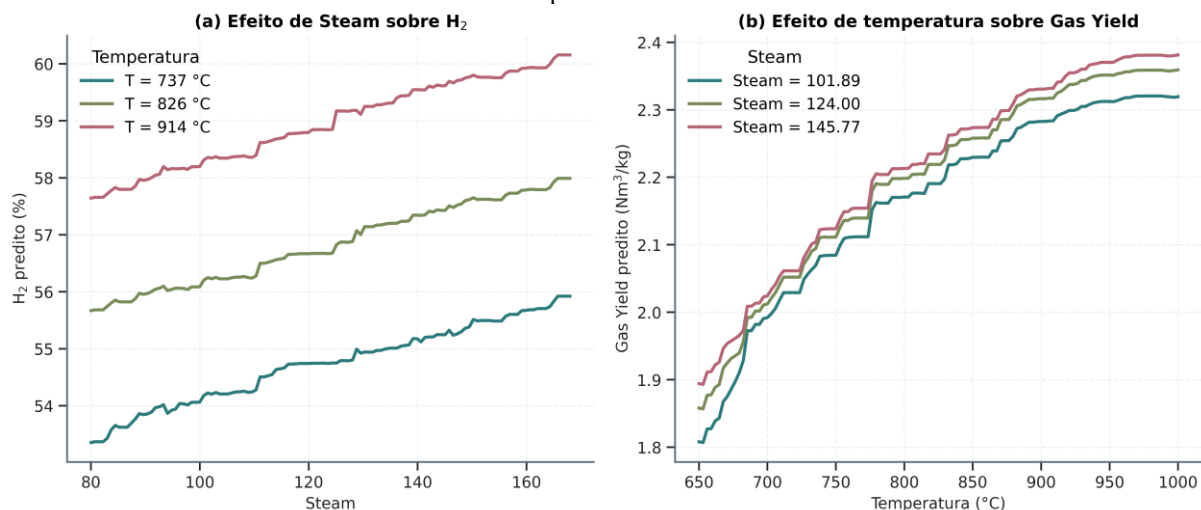
entrada foram mantidas fixas em seus valores medianos, de modo a evidenciar tendências representativas do processo.

Conforme mostrado na Figura 3(a), a fração molar de hidrogênio aumenta de forma monotônica com o incremento da vazão de vapor, sendo esse efeito mais pronunciado em temperaturas mais elevadas. Esse comportamento é consistente com a intensificação das reações de reforma a vapor e deslocamento gás-água (water-gas shift), que favorecem a formação de H_2 na presença de maior quantidade de vapor. Além disso, o aumento da temperatura potencializa essas reações endotérmicas, resultando em níveis mais elevados de hidrogênio no gás de síntese. Tendências semelhantes são amplamente reportadas na literatura para processos de gaseificação com vapor (Lu et al., 2014).

Na Figura 3(b), observa-se que a produção de gás (Gas Yield) aumenta com a elevação da temperatura, independentemente do nível de vapor considerado. Esse comportamento está associado ao favorecimento de reações endotérmicas de gaseificação e à maior conversão da fase sólida em produtos gasosos em condições de maior severidade térmica. O efeito da vazão de vapor sobre o rendimento gasoso também é positivo, embora menos pronunciado do que o efeito da temperatura, indicando que, na faixa operacional avaliada, a temperatura é a variável dominante para a maximização da produção de gás. Esse resultado está de acordo com estudos que destacam o papel central da temperatura no controle da conversão e do rendimento em processos de gaseificação (Yalçinkaya et al., 2016).

De forma geral, os resultados indicam que tanto a fração de hidrogênio quanto a produção de gás são favorecidas por condições de maior severidade operacional, caracterizadas por altas temperaturas e maiores vazões de vapor. No entanto, observa-se que o aumento de H_2 é mais sensível à variação conjunta dessas duas variáveis, enquanto o Gas Yield responde predominantemente à temperatura. Essa diferença reforça que estratégias de otimização do processo devem considerar objetivos distintos, uma vez que condições ótimas para maximização de hidrogênio nem sempre coincidem exatamente com aquelas que maximizam o rendimento gasoso.

Figura 3. Análise de cenários baseada no modelo XGBoost, ilustrando (a) o efeito da vazão de vapor (Steam em kg/h) sobre a fração molar de hidrogênio (H_2) em diferentes temperaturas e (b) o efeito da temperatura sobre a produção de gás (Gas Yield) para diferentes níveis de vapor. As demais variáveis de entrada foram mantidas fixas em seus valores medianos, permitindo a avaliação isolada dos efeitos das condições operacionais.



4. Conclusão

Este trabalho apresentou o desenvolvimento de modelos baseados em XGBoost para a predição da fração molar de hidrogênio e da produção de gás em processos de co-gaseificação biomassa-plástico com vapor. Os resultados demonstraram que o modelo é capaz de capturar adequadamente as relações entre composição da alimentação, condições operacionais e respostas do processo, com bom desempenho preditivo mesmo sob uma estratégia de validação mais rigorosa baseada em particionamento por clusters.

A análise das variáveis revelou que a produção de hidrogênio é fortemente influenciada pelo teor de hidrogênio da alimentação e pela temperatura, enquanto a produção de gás é predominantemente governada pela composição elementar, especialmente pelos teores de carbono e oxigênio. Adicionalmente, a análise de cenários

confirmou que o aumento da temperatura e da vazão de vapor favorece ambas as respostas, embora com diferentes níveis de sensibilidade.

Por fim, destaca-se que as tendências observadas foram obtidas a partir de um modelo treinado em um domínio amplo e fisicamente consistente, o que confere robustez às interpretações apresentadas e reforça o potencial do modelo como ferramenta para análise exploratória e suporte à tomada de decisão em processos de gaseificação com vapor.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os autores agradecem à Fundação de Apoio à Pesquisa do Estado de Minas Gerais - FAPEMIG pelo apoio financeiro.

Referências

- Ascher, S., Wang, X., Watson, I., Sloan, W., & You, S. (2022). Interpretable machine learning to model biomass and waste gasification. *Bioresource Technology*, 364. <https://doi.org/10.1016/j.biortech.2022.128062>
- Bernardes, V. H. X., Gedraite, R., Spogis, N., & Arvelos Altino, S. (2026). Integration of PRO/II simulation and XGBoost machine learning for predicting and optimizing hydrogen-rich syngas from air-blown biomass–plastic co-gasification. *International Journal of Hydrogen Energy*, 215. <https://doi.org/10.1016/j.ijhydene.2026.153747>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- De Las Obras-Loscertales, M., Rufas, A., De Diego, L. F., García-Labiano, F., Gayán, P., Abad, A., & Adánez, J. (2015). Morphological analysis of sulfated Ca-based sorbents under conditions corresponding to oxy-fuel fluidized bed combustion. *Fuel*, 162, 264–270. <https://doi.org/10.1016/j.fuel.2015.09.016>
- Janajreh, I., Adeyemi, I., Raza, S. S., & Ghenai, C. (2021). A review of recent developments and future prospects in gasification systems and their modeling. *Renewable and Sustainable Energy Reviews*, 138. <https://doi.org/10.1016/j.rser.2020.110505>
- Lu, S., Feng, W., Kong, X., & Wu, Y. (2014). Analysis and case studies of residential heat metering and energy-efficiency retrofits in China's northern heating region. *Renewable and Sustainable Energy Reviews*, 38, 765–774. <https://doi.org/10.1016/j.rser.2014.07.015>
- Luo, Y., & Chen, J. (2024). Hydrogen production from biomass steam gasification: Experiment and simulation. *Chemical Engineering Science*, 292. <https://doi.org/https://doi.org/10.1016/j.ces.2024.120011>
- Song, H., Yang, G., Xue, P., Li, Y., Zou, J., Wang, S., Yang, H., & Chen, H. (2022). Recent development of biomass gasification for H2 rich gas production. *Applications in Energy and Combustion Science*, 10. <https://doi.org/10.1016/j.jaecs.2022.100059>
- Thaha, A. N., Subburamu, K., Maduraimuthu, D., A, K., Vellaichamy, P., & Desikan, R. (2026). Optimization of catalytic steam gasification for hydrogen production: response surface methodology approach. *Clean Energy*, 10(1), 78–93. <https://doi.org/10.1093/ce/zkaf062>
- Yalçinkaya, B. H., Erikli, Ş., Özilgen, B. A., Olcay, A. B., Sorgüven, E., & Özilgen, M. (2016). Thermodynamic analysis of the squid mantle muscles and giant axon during slow swimming and jet escape propulsion. *Energy*, 102, 537–549. <https://doi.org/10.1016/j.energy.2016.02.077>